Financial Incentives in Health Care Reform: Evaluating Payment Reform in Accountable Care Organizations and Competitive Bidding in Medicare

Dissertation Advisor: Michael E. Chernew, Ph.D. Zirui Song

Financial Incentives in Health Care Reform: Evaluating Payment Reform in Accountable Care Organizations and Competitive Bidding in Medicare

Abstract

Amidst mounting federal debt, slowing the growth of health care spending is one of the nation's top domestic priorities. This dissertation evaluates three current policy ideas: (1) global payment within an accountable care contracting model, (2) physician fee cuts, and (3) expanding the role of competitive bidding in Medicare.

Chapter one studies the effect of global payment and pay-for-performance on health care spending and quality in accountable care organizations. I evaluate the Blue Cross Blue Shield of Massachusetts Alternative Quality Contract (AQC), which was implemented in 2009 with seven provider organizations comprising 380,000 enrollees. Using claims and quality data in a quasi-experimental difference-in-differences design, I find that the AQC was associated with a 1.9 percent reduction in medical spending and modest improvements in quality of chronic care management and pediatric care in year one.

Chapter two studies Medicare's elimination of payments for consultations in the 2010 Medicare Physician Fee Schedule. This targeted fee cut (largely to specialists) was accompanied by a fee increase for office visits (billed more often by primary care physicians). Using claims data for 2.2 million Medicare beneficiaries, I test for discontinuities in spending, volume, and coding of outpatient physician encounters with an interrupted time series design. I find that spending on physician encounters increased 6

percent after the policy, largely due to a coding effect and higher office visit fees. Slightly more than half of the increase was accounted for by primary care physician visits, with the rest by specialist visits.

Chapter three examines competitive bidding, which is at the center of several proposals to reform Medicare into a premium support program. In competitive bidding, private plans submit prices (bids) they are willing to accept to insure a Medicare beneficiary. In perfect competition, plans bid costs and thus bids are insensitive to the benchmark. Under imperfect competition, bids may move with the benchmark. I study the effect of benchmark changes on plan bids using Medicare Advantage data in a longitudinal market-level model. I find that a $1 increase in the benchmark leads to about a $0.50 increase in bids among Medicare managed care plans.

Table of Contents

Chapter 3. Competitive bidding in Medicare Advantage: Effect of benchmark changes on plan bids

Acknowledgements

This dissertation was made possible by the incredible support and guidance of the faculty members on my dissertation committee. They have inspired me, challenged me, and created a most rewarding and enjoyable experience in the Ph.D. program.

I am foremost indebted to Professor Michael Chernew for giving me the opportunity to embark on this series of research projects, for his daily collaboration and feedback, and for his broader mentorship. I am incredibly grateful to Professor John Ayanian for many discussions of research ideas during my early years in the program, for his clinical expertise, and for serving as a physician role model to me. I deeply appreciate Professor Thomas McGuire in numerous ways, from helping me think through countless research ideas in their nascent stages, to always giving his time to debrief after a research talk, and to playing basketball and other fun times outside of the office. I am eternally grateful to Professor Joseph Newhouse for admitting me into the program, for serving as the faculty advisor to my student fellowship from the National Institute on Aging, and for always giving his time and thought to my work.

I have also benefited from the advice, mentorship, and collaboration of other faculty members in the Ph.D. program and broader health policy community, including Professors Katherine Baicker, Amitabh Chandra, David Cutler, Randy Ellis, John Hsu, Haiden Huskamp, Bruce Landon, Frank Levy, Barbara McNeil, Michael McWilliams, Dana Safran, and Jonathan Skinner. My dissertation work took place at the Department of Health Care Policy, where I was the beneficiary of help from many staff members and administrators, including Emily Corcoran, Debby Collins, Wilma Stahura, and the team at the HCP Help Desk.

I thank my fellow classmates in the Ph.D. program, who have provided an amazing community of friends and colleagues. I especially appreciate everyone in the entering cohort of 2008, whose friendship and support have helped make these years meaningful and memorable. They are Mike Botta, Rebecca Cadigan, Laura Garabedian, Keren Ladin, Nick Menzies, Jessica Perkins, Tisamarie Sherry, Sae Takada, Melissa Valentine, and Beth Wikler. Classmates from adjacent cohorts in the Economics track of the Ph.D. program have also provided a wonderful community for exchanging ideas and feedback. I would like to especially acknowledge Martin Andersen, Sebastian Bauhoff, Abby Friedman, Daria Pelech, Sam Richardson, Aaron Schwartz, Tisamarie Sherry, and Jacob Wallace. Our Health Policy program office has always provided tremendous support and been the backbone of the program in many ways. Special thanks goes to Debbie Whitney, Ayres Heller, Kristin Collins, and Joan Curhan.

Generous funding for graduate school and parts of medical school was provided by a National Institute on Aging Predoctoral M.D./Ph.D. National Research Service Award (F30 AG039175), a Pre-doctoral Fellowship in Aging and Health Economics from the National Bureau of Economic Research (T32 AG000186), and a training grant from the Agency for Healthcare Research and Quality to our Health Policy program.

My deepest gratitude goes to my parents, for teaching me the most important lessons in life and never wavering from their unconditional love and support. They have sacrificed so much over so many years for me. I dedicate this dissertation to them. Finally, I am also indebted to my most supportive friend and partner in life in recent years, Katherine Koh, whose undying faith in me and steadfast encouragement have blessed me with a source of strength that always restores me.

Chapter 1

Effect of the Blue Cross Blue Shield of Massachusetts Alternative Quality Contract on health care spending and quality*

* A version of this chapter was previously published as the following:

Zirui Song, Dana G. Safran, Bruce E. Landon, Yulei He, Randall P. Ellis, Robert E. Mechanic, Matthew P. Day, and Michael E. Chernew, "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract," New England Journal of Medicine. 2011 Sept 8;365(10):909-18. Epub 2011 July 13.

1.1. Introduction

The growth of health care spending is a major concern for households, businesses, and state and federal policymakers.[1-3] In response to continued spending growth in Massachusetts following health care reform, Blue Cross Blue Shield of Massachusetts (BCBS), the state's largest commercial payer, implemented the Alternative Quality Contract (AQC) in January of 2009.[4] The AQC is a contracting model based on global payment and pay-for-performance. It is similar to the two-sided accountable care organization (ACO) model specified by the Centers for Medicare and Medicaid Services (CMS) in its proposed ACO regulations.[5]

Global payment has received attention as an alternate financing mechanism to fee-for-service (FFS) because of its ability to control total spending.[6-7] In July of 2009, a Massachusetts state commission voted unanimously to move the state towards global payment within 5 years.[8] In contrast to a one-sided, "shared savings" model in which providers do not bear risk, providers in a global payment model share in savings if spending is below a pre-specified budget, but are also accountable for deficits if spending exceeds the budget. [9-11] This "downside" risk is a stronger tool for spending control.[12-14]

BCBS implemented the AQC in its health maintenance organization (HMO) and point-of-service (POS) enrollee population. These plans require enrollees to designate a primary care physician (PCP), a feature also found in many patient-centered medical home models.[15-19] Presently, the AQC does not extend to PPO enrollees, as they are not required to designate a PCP. Thus, when a provider organization enters the AQC, only its HMO/POS patients are encompassed by the contract.

The AQC contains 3 main features that distinguish it from traditional fee-for-service contracts and from capitation contracts locally and nationally.[4] First, physicians groups, in some cases together with a hospital, enter into 5-year global budget contracts (in contrast to 1-year contracts). Baseline budgets and future budget growth are based on negotiations with BCBS, but no group was given a 2009 budget below its 2008 spending. The budget covers the entire continuum of care, including inpatient, outpatient, rehabilitation, prescription drugs, and long-term care. The PCP's organization is accountable for all enrollee services, regardless of whether the enrollee receives care from her PCP, the PCP's organization, or any other provider. Since the model currently applies to only HMO/POS enrollees, enrollees must seek referrals for specialist care, consistent with those products' benefit designs. During the year, BCBS pays claims on a FFS basis according to negotiated rates, with year-end budget reconciliation.

Second, AQC groups are eligible for pay-for-performance bonuses up to 10 percent of their budget, with ambulatory and hospital measures each comprising half of the bonus (Section 1.6.1). This potential bonus is substantially larger than typical pay-for-performance programs in the US. BCBS sets a range of performance thresholds, or "gates," for each measure at the beginning of the contract which remain fixed throughout the contract.[4] Each measure receives an annual score based on performance. Scores are weighted and aggregated to calculate the bonus amount paid to the AQC group.

Third, AQC groups receive technical support, including spending, utilization, and quality reports from BCBS to assist them in managing their budget and improving quality. In 2009, 7 physician organizations comprising 321 PCP practices and over 4000 total physicians began assuming risk under the AQC for over 25 percent of BCBS

HMO/POS enrollees. Groups ranged from large physician-hospital organizations to small independent practices united by common leadership. Some AQC groups had prior risk contracts from BCBS, while others entered from FFS contracts without financial risk. By 2011, the AQC has grown to 12 groups accounting for 44 percent of HMO/POS enrollees. I evaluate the impact of the AQC on health care spending and ambulatory quality measures in 2009.

1.2. Methods

1.2.1 Population

The population included BCBS enrollees from January, 2006 through December, 2009. From 2,335,593 total HMO and POS members, I excluded 701,079 who were not continuously enrolled for at least one calendar year. The remaining 1,634,514 members comprised the sample for the main analyses. All AQC and non-AQC providers with BCBS patients were included.

1.2.2 Study Design

I used a pre-post, intervention-control, difference-in-difference approach to isolate the AQC effect. For the spending analyses, the pre-intervention period was 2006 through 2008 and post-intervention was 2009. The intervention group consisted of all enrollees who designated PCPs in practices that began assuming risk under the AQC in 2009. Within the intervention group, I also pre-specified 2 subgroups: one consisting of providers who had prior experience with risk-based contracts from BCBS ("prior-risk"), and the other of providers who entered the AQC without BCBS risk experience ("no-prior-risk"). I hypothesized that the AQC would have a bigger effect on spending in the no-prior-risk subgroup.

I compared spending between all AQC groups and control, prior-risk groups and control, as well as no-prior-risk groups and control. For all 3 comparisons, I decomposed the AQC effect in 4 ways to understand the source of spending differences. First, I decomposed by clinical category using the Berenson-Eggers Type of Service (BETOS) classification version 2009 from CMS.[20] Second, I decomposed the findings by site (inpatient versus outpatient) and type of care (professional versus facility). Third, I examined results by quartile of enrollee health risk.

Lastly, I separated the AQC effect into price and utilization components by repricing claims for each service to its median price across all providers in the study period. Repriced spending differences reflect only utilization differences. Moreover, I examined measures of utilization such as admissions or procedures directly. I further decomposed the spending differences due to price (fees) into two potential explanations: differential fee changes (AQC groups may have received lower fee increases than non-AQC) and referral pattern changes (AQC enrollees may have received more care from lower fee providers). This was done by repricing claims to the median 2009 price for each service *within each practice*.

The quality analysis compared performance on ambulatory process measures between all AQC groups and control using 2007-2009 data. These measures are primary care-oriented and under direct control of the AQC groups. The measures follow Healthcare Effectiveness Data and Information Set (HEDIS) specifications, which are used by most health plans. I analyzed individual measures and aggregate measures for chronic care management, adult preventive care, and pediatric care.

For the spending analysis, the dependent variable was aggregate medical spending per member per quarter (combining BCBS spending and enrollee cost sharing). I excluded prescription drugs from the primary analysis because not all enrollees had prescription drug coverage through BCBS. Spending was computed from claims-level FFS payments made within the global budget. This is an accurate measure of medical spending based on utilization and negotiated FFS prices, but it does not capture the quality bonuses or end-of-year budget reconciliation.

For each ambulatory quality measure, the dependent variable was a dichotomous variable indicating whether the measure was met for an eligible member in a given year. Eligibility is defined by member characteristics and diagnosis; for example, diabetes measures are restricted to members with diabetes.

I controlled for age categories, interactions between age and sex, risk score, and secular trends to correct for differences in individual traits across treatment and control groups. Risk scores were calculated by BCBS from current year diagnoses, claims, and demographic information using the diagnostic-cost-group (DxCG/Verisk Health) score system,[22] similar to the method used by CMS for risk adjustment of payments to Medicare Advantage plans. Higher scores denote lower health status and higher expected spending. The DxCG score is calculated from statistical analyses using a national claims database to relate current year spending to current year diagnoses and demographic information.

1.2.3 Empirical Approach

All analyses were conducted at the enrollee-quarter level. I used a one-part generalized linear model with propensity weights,[21] which mitigated differences in individual traits across treatment and control groups. Propensity weights were calculated using age, sex, and risk scores. For the spending analysis, the dependent variable was spending in dollars per member per quarter. The baseline model was not logarithmic-transformed because the risk score is designed to predict dollar spending. Moreover, evidence shows that linear models perform better than more complex functional forms in predicting health spending.[23-26]

I estimated the reduced-form model below, where $\boldsymbol{X}_{it}$ represents the vector of individual characteristics (age categories, interactions between age categories and sex, and the risk score). Additional independent variables included an indicator for the intervention (*aqc*), indicators for each quarter ($\boldsymbol{q}$), and quarter-intervention interactions ($\boldsymbol{q}$**aqc*). I also included an indicator for the post-intervention period (*post*) as well as its interaction with the intervention group (*post*aqc*), which produced the estimate of the policy effect. Huber-White corrections were used to adjust standard errors for clustering of multiple observations for each individual.[27-28]

$$Spending_{it} = \alpha_{it} + \boldsymbol{X}_{it}\delta + \beta_1 post_t + \beta_2 aqc_i + \beta_3(post*aqc)_{it} + \boldsymbol{q}_t\gamma_1 + (\boldsymbol{q}_t*aqc_i)\gamma_2 + \varepsilon_{it}$$

To assess the AQC impact on quality, I estimated an analogous difference-in-difference model. For aggregate quality analysis, I pooled measures and adjusted for measure-level fixed effects. Independent variables were analogous to the spending model, with year indicators in place of quarter indicators.

$$Quality_{it} = \alpha_{it} + \boldsymbol{X}_{it}\delta + \beta_1 post_t + \beta_2 aqc_i + \beta_3 (post*aqc)_{it} + \boldsymbol{yr}_t\lambda_1 + (\boldsymbol{yr}_t*aqc)\lambda_2 + \mu_{it}$$

I conducted a number of sensitivity analyses (Section 1.6.3), including alterations to the statistical model, repeating analysis on subjects who were continuously enrolled throughout the 4-year study period, coding the risk score in deciles, omitting the propensity score weights, and explicitly including enrollee cost sharing in spending. For the quality analysis, I tested robustness of the linear probability model using a logit model. I also tested for risk score changes that would be consistent with the possibility that under a global payment system, physicians may upcode to garner increased payments, which would make AQC patients seem sicker and thus spending adjusted for health status seem lower. This was an issue in the evaluation of Medicare's Physician Group Practice Demonstration.[30] All analyses used STATA software, version 11. Results are reported with 2-tailed P values. The Harvard Medical School Office for Research Subject Protection approved the study.

1.3. Results

1.3.1. Spending

There were 380,142 subjects with at least one year of continuous enrollment from 2006 through 2009 in the intervention group and 1,351,446 such subjects in the control group. Average age, sex distribution, and health risk scores were similar between the groups (Table 1.1).

Health care spending increased for both AQC and non-AQC enrollees in 2009, but the increase was smaller for AQC enrollees (Table 1.2). Statistical estimates indicated that relative to control, the AQC was associated with a $15.51 decrease in average quarterly spending per enrollee in 2009 (p=0.007, 95 percent confidence interval -$27.21 to -$3.81), a 1.9 percent savings. In the models, the interaction of the secular trend with the AQC indicator demonstrated no significant spending trend differences between AQC and non-AQC groups prior to the intervention ($0.89, p=0.28).

Procedures, imaging, and tests accounted for over 80 percent of the savings. Further decomposition showed that savings largely derived from spending on facility services in the outpatient setting. There were no significant changes in inpatient spending or spending on physician services. The decomposition by member health status showed that members in the highest risk quartile accounted for most (-$14.75, p=0.01) of the savings (Figure 1.1).

Models with standardized prices revealed no AQC effect on utilization. This was supported by quantity analyses of procedures, imaging, tests, admissions, and office visits. Thus, the observed savings reflect differences in price (Section 1.6.4).

Table 1.1. Characteristics of the Study Population.*

Variable	**All AQC Groups** (N=380,142)		**Control Group** (N=1,351,446)	
	Pre-AQC (2006-08)	Post-AQC (2009)	Pre-AQC (2006-08)	Post-AQC (2009)
Member characteristics				
Age — yr	34.4 ± 18.6	35.3 ± 18.5	35.3 ± 18.7	35.5 ± 18.8
Female sex — %	52.6	51.2	51.8	51.0
Health risk score[&]	1.08 (0.12—1.29)	1.16 (0.13—1.39)	1.11 (0.11—1.33)	1.16 (0.12—1.39)

* Plus-minus values are mean ± SD. Values in parentheses are the 25th and 75th percentiles. The number of total enrollees by summing treatment and control exceeds 1,634,514 because of enrollees who had a PCP in the treatment group and another in the control group for at least one year in each case.

[&] Health risk score denotes enrollee health status and expected spending. It is calculated using the DxCG/Verisk Health system, which uses statistical analyses based on a national claims database to relate current year spending to current year diagnoses and demographic information. The DxCG method is a commonly used, proprietary method similar to Medicare's Hierarchical Condition Category (HCC) system used for risk adjustment of prospective payments to Medicare Advantage plans (and developed by the same organization). DxCGs are designed for the under-65 population and are more detailed than the HCC system. Among all subjects, it ranged from 0 to 66 (mean = 1.13, standard deviation = 1.86).

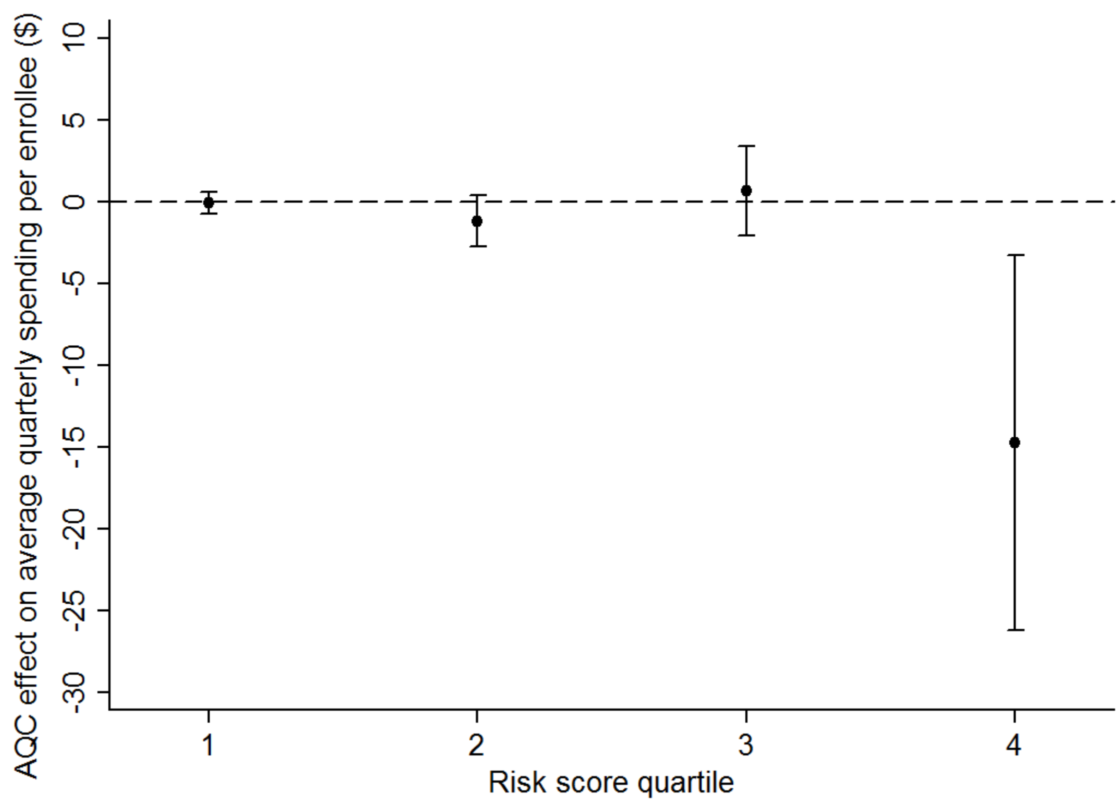
Table 1.2. Change in average health care spending per member per quarter in AQC and non-AQC groups.*

A. All AQC groups vs. control	**All AQC Groups** (N=380,142)			**Control Group** (N=1,351,446)			**Between-Group Difference (p-value)**	
	Pre	Post	Change	Pre	Post	Change		
Total quarterly spending ($)	756	808	53	785	854	69	-15.51	(0.009)
Spending by BETOS category †								
1. Evaluation and management	180	206	25	181	208	27	-2.22	(0.002)
2. Procedures	166	176	10	168	184	16	-5.96	(0.001)
3. Imaging	94	102	8	100	112	11	-3.47	(<0.001)
4. Test	67	75	7	74	85	11	-3.72	(<0.001)
5. Durable medical equipment	10	12	2	11	13	2	-0.14	(0.68)
6. Other	48	50	2	54	55	1	0.80	(0.72)
7. Exceptions/Unclassified	190	189	-1	197	197	0	-0.80	(0.84)
Spending by site and type of care								
Inpatient - professional	35	36	2	34	37	2	-0.72	(0.38)
Inpatient - facility	152	154	3	156	158	3	0.23	(0.95)
Outpatient - professional	316	350	34	300	334	34	-0.28	(0.80)
Outpatient - facility	214	230	16	255	285	30	-14.50	(<0.001)
Ancillary	39	39	-1	40	40	0	-0.24	(0.86)

The intervention group comprised enrollees whose primary care physicians were in the Alternative Quality Contract (AQC) system of Blue Cross Blue Shield of Massachusetts, and the control group comprised enrollees whose primary care physicians were not part of the AQC system. Blue Cross Blue Shield of Massachusetts implemented the AQC system in 2009. All spending figures are in 2009 U.S. dollars.

† The clinical categories were designated according to the Berenson–Eggers Type of Service (BETOS) classification, version 2009.

Figure 1.1. Difference-in-difference estimates of the AQC effect on average health care spending per member per quarter, by risk quartile (all AQC groups versus control).*



* Point estimates with 95 percent confidence intervals from the baseline model.

This price effect could arise either because AQC groups received lower fee increases or because AQC enrollees were shifted to lower-fee providers. I found no differences in price trends (including hospital, non-hospital facility, and physician fees) for AQC and non-AQC providers. The model with prices standardized by physician practice revealed that the price effect was due to referral pattern changes whereby AQC patients were referred to lower-fee providers. Those providers could be in non-hospital settings (such as ambulatory surgery centers) or simply be hospitals that had lower negotiated fees for outpatient care than other hospitals. This model demonstrated that referral shifts accounted for over 90 percent of the AQC-associated relative decrease in quarterly spending (-$14.21, p<0.001) in 2009 (Section 1.6.4).

The prior-risk subgroup compared to control incurred statistically insignificant total savings of $9.29 (p=0.13, -$21.45 to $2.86), or 1.1 percent, per member per quarter. In contrast, the no-prior-risk subgroup experienced larger and statistically significant savings of $45.52 (p=0.006, -$78.13 to -$12.90) or 6.3 percent, suggesting that it drove the main findings. Subgroup decompositions mirrored decompositions of main findings (Section 1.6.2). An interaction test of the differential AQC effect between the two subgroups yielded -$32.94 (p=0.06, -$66.72 to $0.83). Sensitivity analyses supported these results (Section 1.6.3).

1.3.2. Quality

The AQC was associated with a 2.6 percentage-point increase in eligible members meeting quality thresholds for chronic care management (p<0.001) and a 0.7 percentage-point increase in eligible members meeting pediatric care thresholds (p=0.001) (Table 1.3). The AQC was not associated with significant improvement in adult preventive care. Comparisons between prior-risk and control as well as no-prior-risk and control yielded similar results (not shown).

Table 1.3. Change in ambulatory quality performance in AQC and non-AQC groups.*

Quality metric (% of eligible enrollees who met the performance threshold)	**All AQC groups**			**Control groups**			**Difference** (percentage points)		
	Pre	Post	Change	Pre	Post	Change	Unadj.	Adjusted	(P value)
Chronic Care Management (aggregate)	79.1	82.4	3.3	79.6	80.1	0.5	2.8	2.6	(<0.001)
Cardiovascular LDL screening	88.6	90.4	1.8	90.2	90.3	0.1	1.7	1.8	(0.04)
Diabetes: HbA1c testing	89.3	92.0	2.7	89.3	90.2	0.9	1.8	1.7	(<0.001)
Diabetes: eye exam	58.5	63.6	5.1	61.3	60.8	-0.5	5.6	5.5	(<0.001)
Diabetes: LDL screening	86.6	90.5	3.9	86.3	87.3	1.0	2.9	2.8	(<0.001)
Diabetes: Nephrology screening	85.1	87.4	2.3	83.5	84.2	0.7	1.6	1.6	(0.001)
Depression: acute Rx	67.2	66.4	-0.8	66.9	66.9	0.0	-0.8	-1.1	(0.59)
Depression: continuation Rx	51.2	52.0	0.8	50.9	50.2	-0.7	1.5	1.1	(0.59)
Adult Preventive Care (aggregate)	75.7	79.3	3.6	72.8	76.2	3.4	0.2	0.1	(0.67)
Breast cancer screening	80.2	83.2	3.0	79.5	81.9	2.4	0.6	0.6	(0.006)
Cervical cancer screening	87.3	87.6	0.3	84.4	85.2	0.8	-0.5	-0.5	(0.002)
Colorectal cancer screening	64.2	70.7	6.5	60.0	66.5	6.5	0.0	0.0	(0.97)
Chlamydia screening (ages 21-24)	58.6	64.5	5.9	53.4	60.1	6.7	-0.8	-0.8	(0.41)
No antibiotic: acute bronchitis[&]	18.7	25.9	7.2	19.5	21.1	1.6	5.6	5.5	(<0.001)
Pediatric Care (aggregate)	79.5	81.8	2.3	74.6	76.6	2.0	0.3	0.7	(0.001)
Appropriate testing for pharyngitis	93.9	96.0	2.1	82.1	88.4	6.3	-4.2	-3.9	(<0.001)
Chlamydia screening (ages 16-20)	54.8	63.7	8.9	51.1	54.7	3.6	5.3	5.4	(<0.001)
No antibiotic: upper respiratory infection	94.9	95.8	0.9	91.6	92.8	1.2	-0.3	-0.4	(0.52)
Well care: baby (ages <15 mo.)	93.0	93.1	0.1	92.7	92.9	0.2	-0.1	-0.1	(0.91)
Well care: child (ages 3-6)	92.3	94.1	1.8	90.0	91.2	1.2	0.6	0.6	(0.09)
Well care: adolescent	73.8	76.8	3.0	69.1	71.4	2.3	0.7	0.9	(<0.001)

* Adjusted results are from a propensity-weighted difference-in-difference model controlling for all covariates and secular trends. The three aggregate analyses used pooled observations and are further adjusted for measure fixed effects.

1.3.3. BCBS Payments

The AQC-associated savings do not imply that total payments made by BCBS declined. Total BCBS payments must take into account quality bonuses and end-of-year budget surpluses paid to the AQC groups. In 2009, quality bonuses were generally between 3-6 percent of budgets. Additional BCBS support for information technology, staffing, and other needs were between 0-2 percent of budgets. Moreover, all AQC groups spent under their 2009 budget targets, earning on average 3 percent in budget surpluses (consistent with the findings). Taken together, these first year investments and payouts exceeded the average estimated savings of 1.9 percent, suggesting BCBS total payments rose for AQC groups in the first year.

1.4. Discussion

The AQC was associated with modestly lower medical spending and improved quality in the first year after implementation. The savings largely derived from shifting outpatient care to lower-fee providers, and mostly from high risk enrollees. Savings were larger among providers previously paid FFS by BCBS. These results were robust to a series of sensitivity analyses and do not appear to be attributable to upcoding. In addition, spending trends prior to the intervention were not statistically different between the AQC and non-AQC groups.

Quality improvements are likely due to a combination of substantial financial incentives and BCBS data support. AQC quality bonuses are much higher than most pay-for-performance programs in the US, as it applies to the entire global budget rather than to only physician or PCP services.[31]

This study is subject to several limitations. The study population was young and included only POS and HMO members. Thus, results may not generalize to the Medicare population, PPO or indemnity plan enrollees, or other states. However, effects were greater for enrollees who had higher expected spending, so programs serving older populations may experience even larger savings. Furthermore, I do not observe details of each AQC contract, which varied to some degree, or details of provider risk contracting with other payers. While the results suggest quality improved, process measures do not completely capture quality. Formal evaluation of outcome measures could not be conducted due to the lack of pre-AQC enrollee-level outcomes data. However, a weighted average of 5 outcomes metrics at the provider organization level suggests that AQC groups achieved better or comparable outcomes in 2009 compared to recent BCBS network averages (Section 1.6.5).

The findings do not imply that overall spending fell for BCBS in the first year. This reflects the design of the AQC, which focuses on slowing the growth of spending and improving quality initially rather than saving money in the first year. The AQC targets were set based on actuarial projections to save money over the 5-year contract, even after anticipated quality bonuses. Initial investments help to motivate participation and support the delivery system changes required for providers to succeed in managing spending and improving quality. Because provider groups mostly bear the risk, fiscal success from the insurer perspective depends on how well budgets and bonuses are set.

In total, the magnitude of savings was modest. Sustainability of the AQC and the financial viability of the model for providers will ultimately depend on identifying and addressing clinically inefficient care and changing utilization patterns. Nevertheless, the findings on referral pattern changes and quality improvements suggest that provider groups changed behavior in 2009. Importantly, referral pattern changes can subsequently affect pricing in the health care market, as high-price facilities feel pressure from decreased volume. Future studies will need to assess whether changes in utilization and the broader market lead to larger savings.

This initial evaluation offers several lessons for payment reform.[32-34] First, quality need not be threatened by global payment, and process measures can improve given clinically-aligned incentives. Other aspects of quality remain to be evaluated. Second, global payment can introduce greater price competition into the market as referrals move from high-price to low-price facilities. This is a bigger issue for private purchasers since Medicare regulates prices. Finally, even under strong financial incentives, utilization will not change rapidly. Slowing the growth rate of health care spending will ultimately depend on budget updates and the ability of providers to practice in this new environment.

1.5. References

1. Orszag PR, Ellis P. The Challenge of Rising Health Care Costs—A View from the Congressional Budget Office. New Engl J Med 2007;357(18):1973-5.
2. Chernew ME, Baicker K, Hsu J. The Specter of Financial Armageddon — Health Care and Federal Debt in the United States. New Engl J Med 2010;362(13):1166-8.
3. Chernew ME, Hirth RA, Cutler DM. Increased Spending On Health Care: Long-Term Implications for the Nation. Health Aff (Millwood) 2009;28(5):1253-5.
4. Chernew ME, Mechanic RE, Landon BE, Safran DG. Private-Payer Innovation In Massachusetts: The 'Alternative Quality Contract.' Health Aff (Millwood) 2011;30(1):51-61.
5. Centers for Medicare and Medicaid Services. Medicare Shared Savings Program: Accountable Care Organizations. *Federal Register* 2011 Apr 7;76(67):19528-654.
6. Miller H. From Volume to Value: Better Ways to Pay for Health Care. Health Aff (Millwood) 2009;28(5):1418-28.
7. Engleberg Center for Health Care Reform. Bending the Curve: Effective Steps to Address Long-Term Health Care Spending Growth. The Brookings Institution, August 2009. (http://www.brookings.edu/reports/2009/0901_btc.aspx)
8. Bebinger M. Mission Not Yet Accomplished? Massachusetts Contemplates Major Moves on Cost Containment. Health Aff (Millwood) 2009;28(5):1373-81.
9. Health Policy Brief: Accountable Care Organizations. Health Affairs, July 27, 2010. (http://www.healthaffairs.org/healthpolicybriefs/brief_pdfs/healthpolicybrief_20.pdf)

10. Fisher ES, Staiger DO, Bynum JP, Gottlieb DJ. Creating accountable care organizations: The extended hospital medical staff. Health Aff (Millwood) 2007;26(1):w44-57.

11. McClellan M, McKethan AN, Lewis JL, Roski J, Fisher ES. A national strategy to put accountable care into practice. Health Aff (Millwood) 2010;29(5):982-90.

12. Shortell SM, Casalino LP. Implementing qualifications criteria and technical assistance for accountable care organizations. JAMA 2010;303(17):1747-8.

13. Devers K, Berenson R. Can Accountable Care Organizations Improve the Value of Health Care by Solving the Cost and Quality Quandaries? Urban Institute, October, 2009. (http://www.urban.org/url.cfm?ID=411975)

14. Fisher ES, Shortell SM. Accountable care organizations: accountable for what, to whom, and how. JAMA 2010;304(15):1715-6.

15. Goroll AH, Berenson RA, Schoenbaum SC. Fundamental reform of payment for adult primary care: Comprehensive payment for comprehensive care. J Gen Intern Med. 2007;22:410-5.

16. Bodenheimer T, Grumbach K, Berenson RA. A lifeline for primary care. N Engl J Med 2009;360(26):2693-6.

17. Rittenhouse DR, Shortell SM. The patient-centered medical home: will it stand the test of health reform? JAMA 2009;301:2038-40.

18. Merrell K, Berenson RA. Structuring Payment for Medical Homes. Health Aff (Millwood) 2010;29(5):852-8.

19. Kilo CM, Wasson JH. Practice Redesign and the Patient-Centered Medical Home: History, Promises, and Challenges. Health Aff (Millwood) 2010;29(5):773-8.

20. Centers for Medicare and Medicaid Services. Berenson-Eggers Type of Service, 2009. Available at: https://www.cms.gov/HCPCSReleaseCodeSets/20_BETOS.asp.

21. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1): 41-55.

22. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Iezzoni LI, Ingber MJ, Levy JM, and Robst J. Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC model. Health Care Financing Review 2004;25(4):119-41.

23. Zaslavsky AM, Buntin MB. Too much ado about two-part models and transformation? Comparing methods of modeling Medicare expenditures. J Health Econ 2004;23:525-42.

24. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ 2005;24(3):465-88.

25. Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. J Health Econ 2007;26(1):25-48.

26. Ai C, Norton EC. Interaction terms in logit and probit models. Economic Letters 2003;80:123-9.

27. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817-30.

28. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967:221-33.

29. Casella G, Berger RL. Statistical Inference. 2nd ed. Duxbury: Pacific Grove, 2002.

30. Medicare Payment Advisory Commission. Report to the Congress: Improving Incentives in the Medicare Program. 2009 June.
31. Rosenthal MB, Frank RG, Li Z, Epstein AM. Early experience with pay-for-performance: from concept to practice. JAMA. 2005 Oct 12;294(14):1788-93.
32. Shortell SM, Casalino LP, Fisher ES. How The Center For Medicare And Medicaid Innovation Should Test Accountable Care Organizations. Health Aff (Millwood) 2010;29(7):1293-8.
33. Mechanic R, Altman S. Medicare's Opportunity to Encourage Innovation in Health Care Delivery. N Engl J Med 2010;362(9):772-4.
34. Guterman S, Davis K, Stremikis K, Drake H. Innovation in Medicare and Medicaid Will be Central to Health Reform's Success. Health Aff (Millwood) 2010;29(6):1188-93.

1.6. Supporting Materials

Contents

Section 1.6.1.

Table 1.4. Performance measurements in the Alternative Quality Contract

Ambulatory Measures

Measure	Gate 1	Gate 5	Weight
Depression			
1 Acute Phase Prescription	65.3	80.0	1.0
2 Continuation Phase Prescription	49.6	70.0	1.0
Diabetes			
3 HbA1c Testing (2X)	69.9	83.2	1.0
4 Eye Exams	58.0	72.1	1.0
5 Nephropathy Screening	79.7	91.4	1.0
Cholesterol Management			
6 Diabetes LDL-C Screening	85.3	93.8	1.0
7 Cardiovascular LDL-C Screening	85.3	93.8	1.0
Preventive Screening/Treatment			
8 Breast Cancer Screening	77.1	90.0	1.0
9 Cervical Cancer Screening	83.5	92.4	1.0
10 Colorectal Cancer Screening	65.2	83.3	1.0
Chlamydia			
11 Ages 16-20	45.9	63.7	0.5
12 Ages 21-24	50.1	67.3	0.5
Adult Respiratory Testing/Treatment			
13 Acute Bronchitis	*Reporting Only 2009, 2010*		1.0
Medication Management			
14 Digoxin Monitoring	83.9	91.6	1.0
Pedi: Testing/Treatment			
15 Upper Respiratory Infection (URI)	90.6	97.7	1.0
16 Pharyngitis	83.1	99.6	1.0
Pedi: Well-visits			
17 < 15 months	91.8	99.3	1.0
18 3-6 Years	85.5	99.2	1.0
19 Adolescent Well Care Visits	60.0	87.7	1.0
Diabetes			
20HbA1c in Poor Control	45.0	4.7	3.0
21LDL-C Control (<100mg)	33.4	75.6	3.0
22Blood Pressure Control (130/80)	30.9	47.3	3.0
Hypertension			
23Controlling High Blood Pressure	71.6	82.5	3.0
Cardiovascular Disease			
24LDL-C Control (<100mg)	33.4	75.6	3.0
Patient Experiences (C/G CAH PS/ACES)			
25 Communication Quality	91.0	98.0	1.0
26 Knowledge of Patients	80.0	95.0	1.0
27 Integration of Care	80.0	96.0	1.0
28 Access to Care	79.0	96.0	1.0
Patient Experiences (C/G CAH PS/ACES)			
29 Communication Quality	95.0	97.0	1.0
30 Knowledge of Patients	89.0	93.0	1.0
31 Integration of Care	85.0	91.0	1.0
32 Access to Care	70.0	90.0	1.0

Hospital Measures

Measure	Gate 1	Gate 5	Weight
AMI			
1ACE/ARB for LVSD	89.1	98.9	1.0
2Aspirin at arrival	98.3		1.0
3Aspirin at discharge	98.2		1.0
4Beta Blocker at arrival *	96.9		1.0
5Beta Blocker at discharge	98.5		1.0
6Smoking Cessation	93.1	99.9	1.0
Heart Failure			
7 ACE LVSD	87.3	98.9	1.0
8 LVS function Evaluation	95.1	100.0	1.0
9 Discharge instructions	71.4	98.5	1.0
10 Smoking Cessation	88.3	99.6	1.0
Pneumonia			
11 Flu Vaccine	77.8	98.6	1.0
13 Antibiotics w/in 6 hrs	95.6	99.8	1.0
14 Oxygen assessment	100.0		1.0
15 Smoking Cessation	86.7	99.8	1.0
16 Antibiotic selection	87.4	95.4	1.0
17 Blood culture	91.0	98.0	1.0
Surgical Infection			
18 Antibiotic received	86.5	98.9	1.0
19 Received Appropriate Preventive	94.1	99.4	1.0
20 Antibiotic discontinued	77.9	96.2	1.0
21 In-Hospital Mortality - Overall	2.15	0.88	1.0
22 Wound Infection	0.30	0.09	1.0
23 Select Infections due to Medical Care	0.18	0.02	1.0
24 AMI after Major Surgery	0.55	0.10	1.0
25 Pneumonia after Major Surgery	1.57	0.60	1.0
26 Post-Operative PE/DVT	0.93	0.22	1.0
27 Birth Trauma - injury to neonate	0.20	0.01	1.0
28 Obstetrics Trauma-vaginal w/o instrument	3.54	1.54	1.0
Hospital Patient Experience (H-CAHPS)			
29 Communication with Nurses	72.6	81.2	1.0
30 Communication with Doctors	78.1	85.5	1.0
31 Responsiveness of staff	58.4	76.4	1.0
32 Discharge Information	77.7	90.4	1.0

* Each quality measure has designated performance thresholds ranging from Gate 1 (low) to Gate 5 (high), which denote absolute performance based on the percent of eligible members who achieved the measure. Scores for all measures are weighted and summed to a total score. Bonus payment (up to 10% of the global budget) is calculated using a non-linear function of the total score.

Section 1.6.2. Sensitivity analyses methods and results

I conducted several sensitivity analyses. First, I restricted analysis to 548,677 individuals with 48-month continuous enrollment. Second, I redefined the linear risk score into categories bounded by its 10th, 25th, 50th, 75th, 90th, and 95th percentiles and reran the models. Third, I controlled for secular trends using a linear time variable rather than quarterly dummies. Fourth, I repeated all models without propensity score weighting. Fifth, I aggregated spending at the annual to test whether results were sensitive to sample size. Sixth, I repeated analyses on the subpopulation of members with pharmacy benefits and included drug spending within the total. Seventh, I removed enrollee cost-sharing from spending and repeated the models. Linear models perform better than more complex functional forms in predicting health spending while avoiding assumptions concerning heteroskedasticity and distribution of spending required in a non-linear context.[23-25] Linear models also avoid complications in interpreting interaction effects.[26] Lastly, I verified the results using a 2-part model, which consisted of logistic regression of the probability of having any spending followed by linear regression of spending conditional on any spending. Estimated coefficients with smearing adjustments are used to calculate expected spending and the delta method was used to estimate standard errors.[29] For the quality analyses, I tested the robustness of findings to the use of a logit model instead of a linear probability model.

Results of sensitivity analyses support the main findings (Table 1.5). Estimates of the AQC effect on spending remain similar to the baseline case and effect sizes for the no-prior-risk subgroup remain larger than for the prior-risk subgroup. Sensitivity analyses results support the main findings in several ways. Restricting analyses to 48-month continuously enrolled members did not alter the estimates, ameliorating adverse selection concerns. Relative savings by risk category were also unchanged after using quintiles or tertiles. Alternate controls for secular trends did not change the findings. Propensity score weights were used to balance traits between intervention and control groups, and their omission did not impact the results. Different levels of spending aggregation also produced similar results. Omitting cost-sharing from all claims did not change the estimates, suggesting that benefit design was not differentially heterogeneous among treatment and control enrollees. Finally, restricting analyses to members with pharmacy benefits with drug spending included did not change the findings. The descriptive analyses showed that enrollees were similar in age, sex, and risk score across intervention and control groups. I also demonstrated that trends in spending prior to the intervention were not statistically different. Finally, the use of a 2-part model with logarithmic transformation and logistic models to analyze spending and quality did not change the results (not shown).

The analyses of risk trends indicate an increase in risk in the AQC group relative to the controls. This might reflect a true increase in risk, which is important to control for, or upcoding. The finding was much smaller and not statistically significant in the no-prior-risk group, which accounts for the largest savings. Nevertheless, if one assumed all of the increase in risk was due to upcoding, the AQC effect would be attenuated by between 18 to 38 percent.

Although there was an increase in risk among the AQC group, the results are likely not driven by upcoding for several reasons. First, the results are robust to measuring risk in deciles, which would be less sensitive to upcoding. Second, the increased risk was not apparent in the no-prior-risk group, where savings is largest. Third, upcoding would presumably affect the analyses of utilization, and lead me to erroneously conclude that volumes decreased after the intervention. However, I found no AQC effect on utilization. Fourth, upcoding would likely affect both professional and facility spending and fees, but I found effects only on facility spending and fees.

Table 1.5. Sensitivity analyses of difference-in-difference estimates of the association between the AQC and health care spending.*

All AQC vs. Non-AQC	P value	**Prior-risk vs. Non-AQC**	P value	**No-prior-risk vs. Non-AQC**	P value
Association between the AQC and average quarterly spending per enrollee ($)					
Main estimates					
-15.51	0.009	-9.29	0.13	-45.52	0.006
Sensitivity analyses					
(1) Restricting analyses to the subset of 48-month continuous enrolled members					
-18.79	0.01	-12.64	0.11	-42.21	0.050
(2) Controlling for health risk score as categorical rather than continuous variable					
-16.08	0.01	-10.30	0.11	-46.31	0.008
(3) Controlling for secular time trend as continuous rather than categorical variable					
-24.21	0.050	-19.21	0.13	-34.35	0.30
(4) Omitting propensity score weights					
-15.85	0.008	-9.54	0.13	-45.82	0.006
(5) Annual rather than quarterly spending as the dependent variable					
-63.90	0.006	-42.47	0.08	-162.73	0.01
(6) Omitting cost-sharing from claims					
-16.17	0.007	-9.80	0.12	-46.92	0.004
(7) Restricting analyses to members with pharmacy benefits and including drug spending					
-19.18	0.005	-13.14	0.07	-48.02	0.009

* Main estimates are from the baseline difference-in-difference model using quarterly spending as the dependent variable, age categories, age-sex interactions, continuous risk, indicators for each quarter, and quarter*AQC interactions. The baseline model uses the sample of members continuously enrolled for at least 12 months over the 4-year study period and uses propensity weights. Rows (1) through (7) show analogous results from sensitivity analyses. Results of the annual spending model (5) should be divided by 4 to be comparable to the quarterly estimates. All analyses use Huber-White robust standard errors.

Section 1.6.3. Decomposition of spending results into price and utilization effects.

This section details the decomposition of the difference-in-difference estimate of the association between the AQC and health care spending into its price and utilization (quantity) components. To better understand how the AQC groups achieved lower spending relative to non-AQC groups in 2009, I first repriced outpatient facility claims for each service to its median price across all providers in all 4 years of the study period. I performed repriced analysis on only outpatient facility claims because outpatient facility spending accounted for the bulk of the association between AQC and spending (see Table 1.2 and Table 1.6). Moreover, inpatient services are priced differently (diagnostic related groups or DRGs, rather than fees).

Repeating the difference-in-difference regression analyses using the repriced claims allowed me to isolate the share of spending differences due to utilization (quantity), rather than price. I found that, relative to control, the AQC was associated with a $2.62 (p=0.25) decrease in average quarterly spending per enrollee in 2009. Compared to the main estimate of a $15.51 savings associated with the AQC (decrease in spending relative to control), this repriced estimate is small and insignificant, suggesting that the main result is driven by price differences, not by utilization. Sensitivity analyses using direct measures of utilization (quantity) as the dependent variable in the model, such as the number of procedures, images, tests, admissions, or office visits, yielded no negative and significant association between the AQC and utilization. This result, however, still left two potential explanations for the price effect on spending.

Thus, I further decomposed the spending differences into two potential price-driven explanations: differential fee changes (AQC groups may have received lower fee increases than non-AQC) and referral pattern changes (AQC enrollees may have received care from lower fee providers). This was done by repricing claims to median 2009 prices for each service *within each practice*. This differs from the first repricing method above because it retains the variation in prices *between* practices, while standardizing prices *within* each practice (the Massachusetts Office of the Attorney General has documented the variation in prices between practices[*]). The resulting association between AQC and spending reflects changes in the share of spending across practices, which I denoted the referral effect. This difference-in-difference model produced an AQC-associated relative spending decrease of $14.21 ($p<0.001$), which is approximately 90 percent of the main AQC-associated savings in year 1. This finding was supported by further sensitivity analyses. For example, I found no consistent differences in price trends (including hospital, non-hospital facility, and physician fees) for AQC and non-AQC providers.

* See "Examination of Health Care Cost Trends and Cost Drivers: Report for Annual Public Hearing, March 16, 2010" released by the Office of the Attorney General Martha Coakley.

Section 1.6.4.

Table 1.6. Change in average health care spending per member per quarter in AQC and non-AQC groups, prior-risk subgroup vs. control and no-prior-risk subgroup vs. control.*

Prior-risk subgroup vs. control	**Prior Risk AQC groups** (N=341,615)			**Control Group** (N=1,351,446)			**Between-Group Difference (p-value)**	
	Pre	Post	Change	Pre	Post	Change		
Total quarterly spending ($)	757	816	58	781	850	69	-9.29	(0.13)
Spending by BETOS category								
1. Evaluation and management	182	209	27	180	208	27	-0.25	(0.74)
2. Procedures	166	178	12	167	183	16	-4.46	(0.01)
3. Imaging	94	103	8	100	111	11	-2.80	(<0.001)
4. Test	67	75	8	74	85	11	-3.01	(<0.001)
5. Durable medical equipment	10	11	2	11	13	2	-0.16	(0.65)
6. Other	48	50	2	53	54	1	1.03	(0.66)
7. Exceptions/Unclassified	190	190	0	196	196	0	0.36	(0.93)
Spending by site and type of care								
Inpatient - professional	35	37	2	34	36	2	-0.37	(0.66)
Inpatient - facility	152	155	3	154	157	3	0.85	(0.83)
Outpatient - professional	319	355	37	299	333	34	2.85	(0.02)
Outpatient - facility	212	230	18	254	284	30	-12.16	(<0.001)
Ancillary	40	39	-1	40	40	0	-0.46	(0.74)

Table 1.7.

No-prior-risk subgroup vs. control	No Prior Risk AQC groups (N=40,468)			Control Group (N=1,351,446)			Between-Group Difference (p-value)	
	Pre	Post	Change	Pre	Post	Change		
Total quarterly spending ($)	698	725	27	791	859	68	-45.52	(0.006)
Spending by BETOS category								
1. Evaluation and management	164	180	16	181	208	27	-12.38	(<0.001)
2. Procedures	154	158	4	169	185	16	-13.46	(0.003)
3. Imaging	86	92	6	101	112	11	-6.13	(0.001)
4. Test	68	70	2	75	86	11	-9.34	(<0.001)
5. Durable medical equipment	10	12	2	11	13	2	-0.17	(0.86)
6. Other	45	46	1	54	56	1	-0.09	(0.99)
7. Exceptions/Unclassified	171	168	-3	200	199	-1	-3.96	(0.72)
Spending by site and type of care								
Inpatient - professional	30	30	1	34	37	2	-1.81	(0.32)
Inpatient - facility	133	138	5	158	160	2	1.15	(0.91)
Outpatient - professional	275	299	24	301	335	34	-11.27	(<0.001)
Outpatient - facility	227	222	-5	258	287	30	-36.80	(<0.001)
Ancillary	34	37	3	41	40	0	3.21	(0.39)

* Total enrollees in prior-risk and no-prior-risk subgroups exceeds 380,142 because of enrollees who had a PCP in prior-risk and a PCP in no-prior-risk for at least one year in each case. Average spending in the non-AQC group differs slightly across Table 2 and this Section due to adjustments for age, sex, and health risk score across different sample sizes of the treatment group.

Section 1.6.5.

Table 1.8. Unadjusted outcome quality for BCBS and 2009 AQC Groups*

	BCBS Network Average			**AQC Weighted Average**
	2007	2008	2009	2009
Diabetes				
HbA1c Control (<9 percent)	83.7	79.8	82.0	80.7
LDL-C Control (<100 mg/dL)	45.7	51.3	51.3	57.7
Blood Pressure Control (130/80)	30.9	36.7	38.0	44.3
Hypertension				
Controlling High Blood Pressure (140/90)	68.4	70.3	69.5	68.4
Cardiovascular Disease				
LDL-C Control (<100 mg/dL)	64.2	69.5	69.5	69.9

***** Outcome quality scores denote the percent of eligible enrollees for whom the quality performance threshold was met. Scores are weighted by eligible members for each measure. These scores are unadjusted averages.

Chapter 2

Unintended consequences of redistributing Medicare payments from specialists to primary care physicians*

* A version of this chapter has been submitted for publication. Please do not circulate.

2.1. Introduction

Facing an imperative to control the growth of health care spending, policymakers are looking to the physician payment system for savings.[1-4] The Medicare Physician Fee Schedule, which lists the prices that Medicare pays for each physician service under Part B of Medicare, faces multiple impending cuts. The recent failure of the bipartisan "super committee" to agree on a debt reduction plan triggers automatic cuts of 2 percent per year to the fee schedule beginning in 2013. Meanwhile, the sustainable growth rate formula is due to slash Medicare fees by close to 30 percent in 2013.[5] Such across-the-board fee cuts are highly controversial.

As an alternative, some proposals call for lowering specialist fees while preserving or raising primary care physician (PCP) fees. The physician fee schedule contains large payment differences between primary care and specialty services, deriving from the resource-based relative value system.[6,7] These fee differences have been linked to a substantial income gap between PCPs and specialists and the shortage of students entering primary care.[8-11] In 2011, the Medicare Payment Advisory Commission voted to replace the sustainable growth rate system with a 10-year plan that would cut specialist payments in the Medicare fee schedule by 5.9 percent in each of the first 3 years before freezing them for the next 7, while freezing PCP fees for all 10 years.[12] To date, there is little research on the effect of such asymmetric fee cuts.

On January 1, 2010, the Centers for Medicare and Medicaid Services (CMS) undertook a targeted policy to redistribute Medicare payments from specialists to PCPs by cutting specialists fees and raising PCP fees for typical physician visits. Specifically, CMS eliminated payments for consultations from the fee schedule, billed more frequently by specialists, while raising fees for office visits, more frequently used by PCPs.[13]

Similar to consultations, physicians bill for office visits (also known as "evaluation and management") using a set of codes that contains 5 levels of clinical complexity for both "new" patient visits (Healthcare Common Procedure Coding System (HCPCS) codes 99201-99205) and "established" patient visits (99211-99215). Consultations had previously commanded higher payments than office visits at each level of clinical complexity. In 2009 for example, Medicare paid $124.79 on average for a consultation of medium complexity, compared to $91.97 for a new patient office visit and $61.31 for an established patient office visit of medium complexity. In place of consultations, the 2010 policy instructed all physicians to bill for office visits. Despite the simultaneous fee increase for office visits, they remained well below prior consultation fees, with new office visits 16-26 percent lower and established office visits 42-61 percent lower than their consultation counterparts from 2009 (Section 2.6.1). The compensatory increase in office visit fees averaged about 5 percent (Section 2.6.2).

CMS designed the policy to be budget neutral—amounting to a transfer of dollars from specialists to PCPs.[13] Budget neutrality was based on a CMS assumption of no "upcoding."[14] In other words, physicians were assumed to bill for office visits of the same level of complexity as they had done for prior consultation codes. I evaluated the effects of this policy on spending, volume, and intensity of coding for office encounters in the first year of implementation.

2.2. Methods

2.2.1. Study Population

The population consisted of Medicare beneficiaries drawn from the 2007-2010 Thomson Reuters MarketScan Medicare Supplemental claims database. This database consists of a large convenience sample of U.S. Medicare beneficiaries with Medicare Supplemental (Medigap) coverage through large employers. For all of these beneficiaries, Medicare is the primary payer and all Medicare claims are reported in the data.[15]

From a total of 2.9 million beneficiaries, I excluded 326,491 beneficiaries who were not enrolled for at least 1 calendar year. In addition, I excluded 311,271 subjects who were enrolled in health maintenance organization or other plans paid through capitation. Physicians for beneficiaries in capitated plans may not be directly affected by the fee schedule changes. The final analytic sample comprised 2.2 million beneficiaries who were enrolled for at least 1 year. In one of the sensitivity analyses, I further restricted the sample to 798,262 beneficiaries who were continuously enrolled for all 4 years in the data. I compared baseline characteristics of the population before and after the intervention using t-tests for continuous variables and the Pearson Chi-squared test for categorical variables.

2.2.2. Study Design

I used an interrupted time series (segmented regression) design to estimate the effect of the Medicare policy on spending, volume, and intensity of coding for all 3 types of outpatient encounters (consultations, new patient office visits, and established patient office visits). The pre-intervention period was 2007-2009, and the post-intervention period was 2010. I analyzed the policy's effects on all encounters, as well as on each type of encounter.

In the Medicare Physician Fee Schedule, determination of the clinical complexity of a patient-physician encounter is somewhat subjective. For example, a level 1 new patient office visit is characterized by a "problem-focused" history and physical examination involving "straightforward" medical decision-making, typically requiring 10 minutes. In contrast, a level 5 new patient office visit entails a "comprehensive" history and examination involving decision making of "high complexity," typically requiring 60 minutes (Section 2.6.3). I examined changes in the coding of complexity associated with the policy, adjusting for actual health status, baseline trends, and other covariates.

I also decomposed the policy's effect by provider specialty. Specifically, I assigned claims to the PCP category if provider specialty was internal medicine, family medicine, geriatric medicine, pediatrics, or preventive medicine. These providers accounted for about half of the claims. The remaining half was assigned to the specialist category, with a small fraction of non-physicians and facility providers excluded. I used only outpatient claims and included both professional and facility fees for all encounters.

2.2.3. Empirical Approach

For the spending analysis, the dependent variable was spending (in 2010 U.S. dollars). Spending was computed from claims-level total payments, including employer and beneficiary cost sharing. For analyses of volume, the dependent variable was the number encounters. Each encounter was a unique instance of a consultation or office code for a unique beneficiary on a unique day. For the complexity analysis, the dependent variable was the encounter's level of complexity.

Independent variables included age, sex, health risk score, secular trends, indicators for quarter of the year, and indicators for the beneficiary's hospital referral region (HRR). Taking beneficiary demographic information and ICD-9 diagnoses in the claims, I calculated health risk scores using the CMS hierarchical condition categories (CMS-HCC) model, which is used for formal risk adjustment of payments to Medicare Advantage plans.[16] This model generates a risk score based on age, sex, and diagnoses. HRRs are hospital markets constructed based on where patients receive care and are commonly used in the study of geographic variations in spending and utilization.[17,18] Over 90 percent of beneficiaries in Medicare live in one of the 306 HHRs where over 80 percent of residents' care is delivered by providers in that HRR.[19]

Since cost sharing also affects utilization of services, I included each beneficiary's cost sharing as a percentage of total encounter spending in the model as a sensitivity analysis. This is a unique variable in the Marketscan Medicare claims data which is not available in traditional Medicare Part B claims.

All statistical analyses were conducted at the beneficiary-quarter level. I used a multivariate interrupted time series (segmented regression) model of the following form.

$$y_{it} = \alpha_{it} + \boldsymbol{X}_{it}\Pi + \beta_1 trend_t + \beta_2 pre2009_t + \beta_3 post2010_t + \tau_t + \delta_i + \varepsilon_{it}$$

The vector of individual characteristics, $\boldsymbol{X}_{it}$, comprises age categories, interactions between age categories and sex, and the CMS-HCC risk score. I include an underlying linear trend over the 4 year period, a dummy variable for pre-2009 to account for an unrelated policy that increased payments in the 2009 fee schedule, and a dummy variable for post-2010 which identifies the discontinuity in the outcome variable after 2010. The coefficient β_3 reflects the change in average quarterly levels of the outcome after the policy took effect, controlling for pre-intervention trends. In other words, the model captures a discontinuity in “cell” averages in 2010 compared to before, rather than a discrete change at quarter 1 of 2010 from annual trends. The cell average model with an underlying trend is preferred over the year-to-year trend break model for two reasons. First, any volatility in annual trends can dramatically affect the treatment effect given only 4 data points for each year. Second, and more importantly, given only 1 year of post-intervention data, it is likely premature to interpret any significant coefficient on the 2010 trend as a true policy effect. Thus, the model identifies an average level change. Standard errors were clustered at the hospital referral region level, which produces even more conservative standard errors than alternative methods of individual-level generalized estimation equation models.[20,21]

I conducted a number of sensitivity analyses from the baseline model, including omitting quarter fixed effects for seasonality, HRR fixed effects, secular trend, and individual beneficiary characteristics (Section 2.6.4). I also repeated analysis on the continuous 4-year enrollees, as well as repeating analysis after including of cost sharing in the model. All analyses were conducted using STATA software, version 11 (Statacorp; College Station, TX). Results are reported with two-tailed P values. The Harvard Medical School Office for Research Subject Protection approved the study protocol.

2.3. Results

There were 2,247,810 unique Medicare beneficiaries who were continuously enrolled for at least one year during the study period. Table 2.1 describes the age distribution, gender, CMS-HCC risk scores, and region of residence for the population before and after the 2010 policy.

Table 2.1. Characteristics of the Study Population.*

Characteristics	**Medicare Beneficiaries** (N=2,247,810)		
	Before elimination of consultations (2007-09)	After elimination of consultations (2010)	P value
Member characteristics			
Mean age (yrs)	75.0 ± 7.4	75.5 ± 7.4	<0.001
Age distribution (%)			<0.001
65-69 yrs	29.9	26.8	
70-74 yrs	22.2	22.8	
75-79 yrs	20.4	20.4	
80-84 yrs	15.7	16.5	
≥85 yrs	11.7	13.4	
Male sex (%)	45.1	45.5	<0.001
CMS-HCC risk score †	0.56 ± 0.20	0.52 ± 0.21	<0.001
Region of residence (%)			<0.001
Northeast	14.6	14.5	
North central	38.9	44.2	
South	34.3	32.3	
West	12.1	8.9	

* Plus-minus values are mean ± SD. Variables before and after the intervention were compared using t-tests for continuous variables and the Pearson Chi-squared test for categorical variables.

† The CMS-HCC risk score is the concurrent risk score calculated from age, sex, and diagnoses information using the Centers for Medicare and Medicaid Services Hierarchical Condition Category (HCC) system.

2.3.1. Spending

In the first year, the policy eliminated an average of $18.52 per beneficiary per quarter (95% confidence interval [CI], -$19.48 to -$17.55; p<0.001) in consultation spending. At the same time, spending increased by $13.64 (95% CI, $12.89 to $14.40; p<0.001) on new patient office visits and $15.08 (95% CI, $13.56 to $16.60; p<0.001) on established patient office visits per beneficiary per quarter. On net, spending on all physician encounters was higher by $10.20 per beneficiary per quarter after the policy, controlling for pre-existing trends (95% CI, $8.58 to $11.82, p<0.001) (Table 2.2).

These changes represented a more than doubling of spending (131 percent increase) on new patient office visits and a 12 percent increase in spending on established visits. Spending was substantially higher at baseline for established visits than for new visits, confirming that the former comprises the majority of physician’s face-to-face encounters with patients. Net spending on all encounters increased 6 percent after the policy.

Figures 1A and 1B decomposes unadjusted monthly spending per beneficiary by category of physician specialty. Prior to the elimination of consultations, average spending on specialist consultations was three-fold higher than that for PCPs. Correspondingly, spending on new office visits to specialists saw a larger commiserate jump than that for PCPs in 2010 (Figure 2.1A). Spending on established patient visits increased for both PCPs and specialists after the policy (Figure 2.1B), driving an increase overall spending for both groups. The statistical model shows that 58 percent of the increased spending were attributable to PCP encounters, while 42 percent were for specialist encounters (p<0.001).

Table 2.2. Average spending, volume, and coded complexity of outpatient physician encounters before and after 2010.*

	Medicare Beneficiaries (N=2,247,810)				
			Difference (After – Before)		
	Before 2010	After 2010	Unadjusted	Adjusted	P value
Spending ($ per beneficiary per quarter)					
All Encounters	157.04	166.74	9.7	10.20	<0.001
Consultations	19.67	0.30	-19.37	-18.52	<0.001
Office visits (new patient)	10.39	23.67	13.28	13.64	<0.001
Office visits (established patient)	126.98	142.77	15.79	15.08	<0.001
Volume (per 100 beneficiaries per quarter)					
All Encounters	192.4	195.3	2.9	-1.4	0.18
Consultations	11.7	0.2	-11.5	-11.4	<0.001
Office visits (new patient)	9.5	18.2	8.7	8.7	<0.001
Office visits (established patient)	171.2	176.9	5.7	1.3	0.19
Complexity level (1-5)					
All Encounters	3.29	3.34	0.05	0.01	<0.001
Consultations and Office visits (new patient)	3.34	3.40	0.06	0.03	<0.001
Office visits (established patient)	3.29	3.34	0.05	0.01	<0.001

* All subjects are Medicare beneficiaries with Medicare Supplemental coverage. The centers for Medicare and Medicaid Services eliminated payment for consultation codes on January 1, 2010. Spending adjusted to 2010 dollars. Adjusted differences are from a model controlling for age, sex, CMS-HCC risk score, pre-existing trends, seasonality, and HRR fixed effects. Standard errors are clustered by HRR.

Figure 2.1. Spending on outpatient physician encounters*

Panel A.

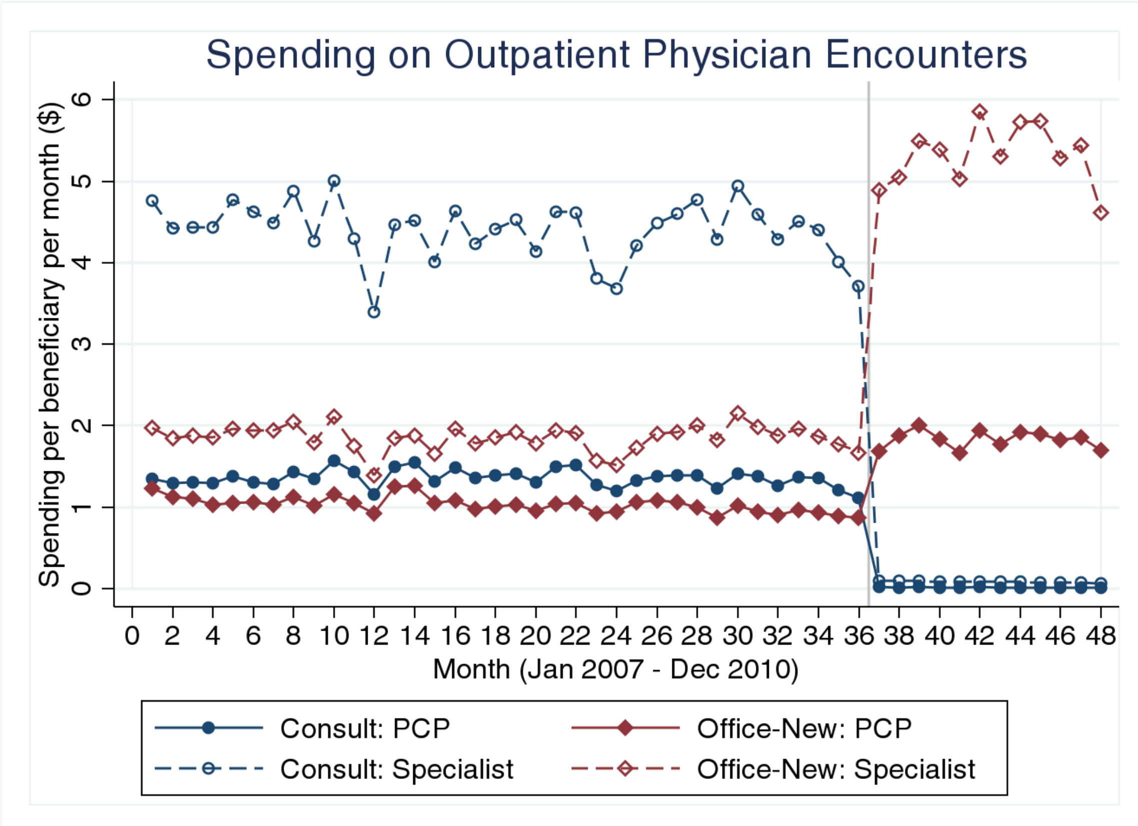


Fig. 2.1 (Continued)

Panel B.

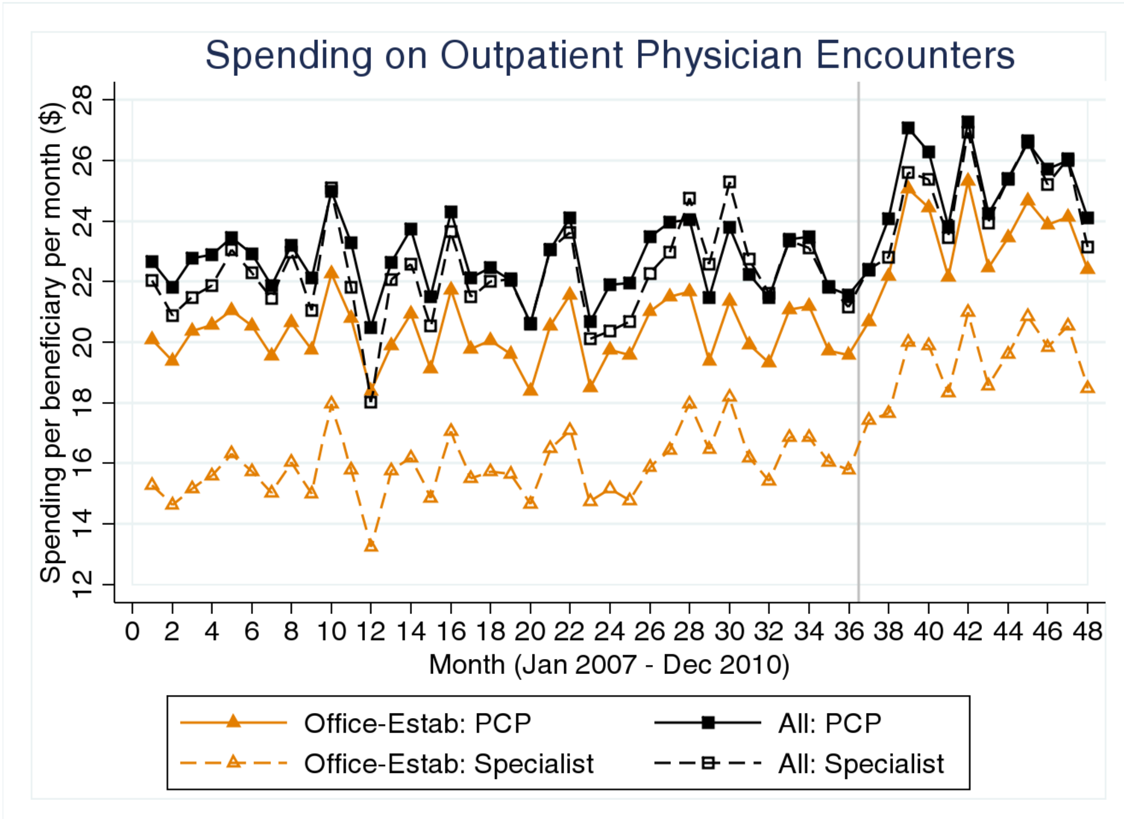


Figure 1 shows spending per beneficiary per month in raw (unadjusted) U.S. 2010 dollars for physician encounters, from January 2007 through December 2010. Each type of encounter is decomposed into primary care physician encounters and specialist encounters. Panel A shows consultations and "new" patient office visits. Panel B contains "established" patient office visits and total spending on all encounters. The centers for Medicare and Medicaid Services (CMS) eliminated payment for consultation codes in January 2010. The gray vertical line is drawn between December 2009 and January 2010 and denotes when the policy took effect.

2.3.2. Volume

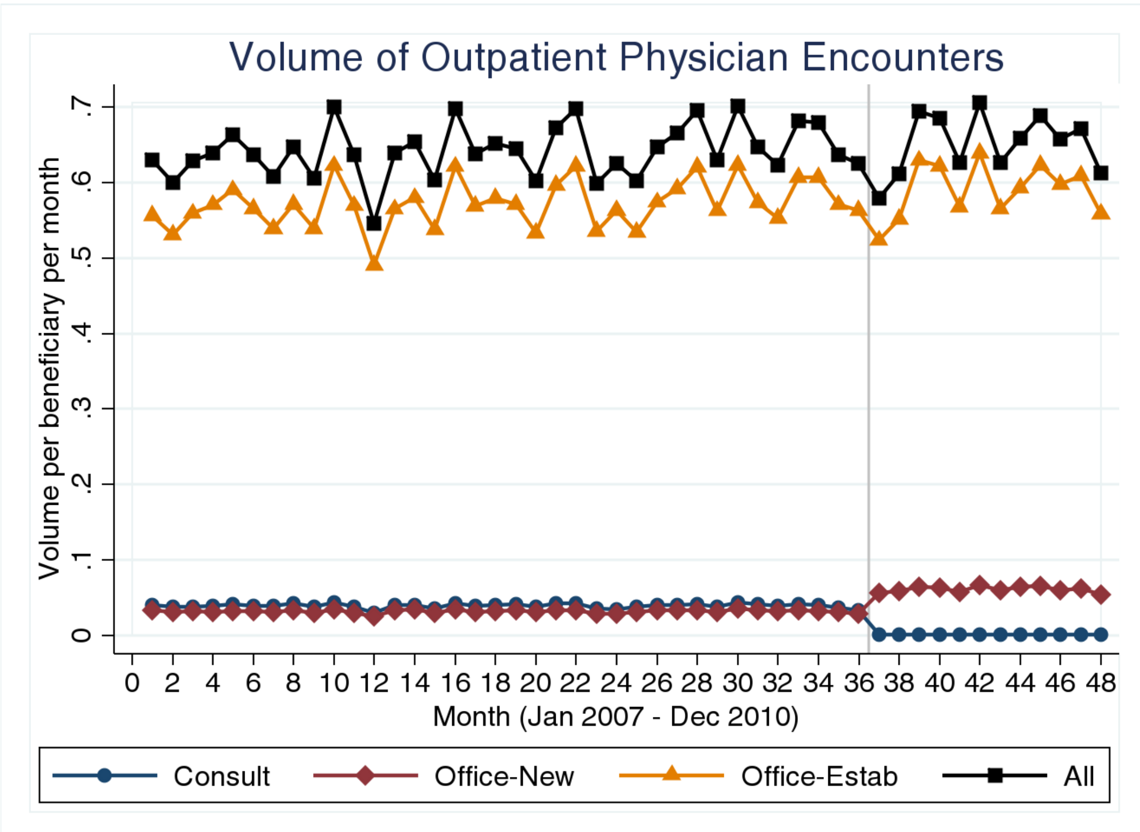
After Medicare's policy took effect, the volume of consultations decreased by 11.4 per 100 beneficiaries per quarter ($p<0.001$). At the same time, the volume of new office visits rose by 8.7 per 100 beneficiaries per quarter ($p<0.001$), substituting for the bulk of consultations that were eliminated (Table 2.2). This was consistent with unadjusted trends (Figure 2). The volume of established office visits increased by a statistically insignificant 1.3 per 100 beneficiaries per quarter ($p=0.19$). On net, there was no significant change in the total volume of encounters (-1.4 per 100 beneficiaries per quarter, $p=0.18$) (Table 2). Subgroup findings for PCPs and specialists were consistent with overall results.

2.3.3. Complexity

Figure 2.3 displays the discrete increase in coded complexity of encounters from before to after January, 2010. Because consultations were largely replaced by new patient office visits, these two encounters were combined in the graph, suggesting an overall increase in the coding level for these visits. In adjusted analyses, the statistical estimates show a discontinuous jump of 0.03 ($p<0.001$) in the average level of complexity for consultations and office visits, as well as an increase of 0.01 ($p<0.001$) in established office visits, after the policy took effect. On net, the magnitude of this coding effect for all encounters was 0.01 ($p<0.001$).

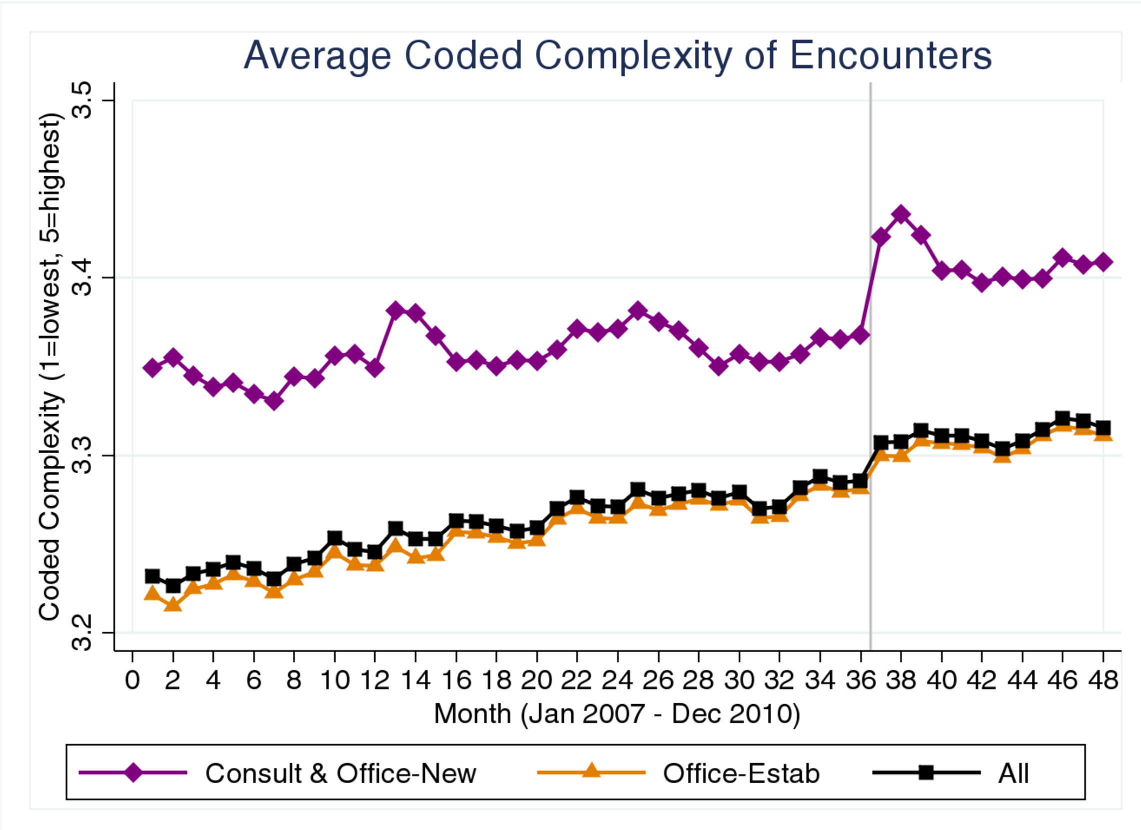
Converting this coding effect to dollars, the model suggests that it was associated with an increase of $5.20 in spending ($p<0.001$). In other words, coding accounted for about 50 percent of the overall increase in spending. Given no significant changes in volume, the results suggest that the other 50 percent of the overall spending effect was attributable to the administrative fee increases (the compensatory fee increase) embedded in the 2010 policy.

Figure 2.2. Volume of outpatient physician encounters*



* Figure 2 shows total encounter volume per beneficiary per month from January 2007 through December 2010. All primary care physician and specialist encounters are included. All 3 types of encounters are drawn on the same scale. The centers for Medicare and Medicaid Services (CMS) eliminated payment for consultation codes in January 2010. The gray vertical line is drawn between December 2009 and January 2010 and denotes when the policy took effect.

Figure 2.3. Average coded complexity of encounters*



* Figure 3 shows the average complexity of encounters from January 2007 through December 2010. Levels of complexity range from 1 to 5. All primary care physician and specialist encounters are included. Consultations and "new" office visits are averaged together as the latter largely substitutes for the former after the policy took effect (Figure 2). The centers for Medicare and Medicaid Services (CMS) eliminated payment for consultation codes in January 2010. The gray vertical line is drawn between December 2009 and January 2010 and denotes when the policy took effect.

Importantly, this coding effect need not reflect “upcoding,” whereby physicians code at a higher than appropriate level for a given service. Given the variety of definitions governing the complexity of encounters (Section 2.6.3), a physician who codes based on time may justifiably determine that in place of a level 3 consultation (40 minutes), an appropriate substitute is a level 4 new patient office visit (45 minutes) rather than a similar level 3 new patient office visit (30 minutes). For an established patient, only a level 5 visit offers the commiserate 40 minutes.

2.4. Discussion

Medicare’s elimination of consultations was associated with increased overall spending of about 6 percent for office encounters in 2010. This increased spending was not explained by changes in volume, but rather by an increased intensity of coding and Medicare’s compensatory fee increase for office visits in 2010. The results suggest that the policy did not achieve its goal of budget neutrality in 2010. Visits to PCPs explained about 58 percent of the overall increase in spending, suggesting that the policy was marginally successful in increasing Medicare payments to PCPs relative to specialists.

There is a dearth of literature examining broad simultaneous fee cuts to specialists and fee increases for PCPs. In contrast to previous literature on Medicare fee cuts,[22-26] this analysis did not demonstrate a volume increase by specialists in response to the fee cut. There are several potential explanations for this finding. First, relative to their fees for procedures and other services, office visit fees for specialists may be low enough that extra volume is not profitable. I imagine this is more likely the case for procedural specialists or surgeons who spend most of their time in the procedural setting, as opposed to cognitive specialists whose portfolio is dominated by office-based care. Any pressure to increase volume may have gone to other, perhaps more profitable, services. I was unable to systematically test for such potential “spillover” effects given other policies that were simultaneously in place in 2010, such as Medicare fee cuts in certain radiology and cardiology services which were directly impacting utilization in those services areas.

Second, volume may have remained unchanged because coding for higher complexity visits (not necessarily outside boundaries of existing definitions) allowed physicians to make up for the fee cut. While this coding effect played a nontrivial role in offsetting the fee cut, the results suggest it was unlikely that physicians made up for the fee cut entirely through coding.

Third, volume may not have responded because most physicians are already at capacity with office visits. Thus, it would be difficult to increase the number of visits within the work day without conceivably working longer or shortening the length of visits. Shorter visits would earn a lower fee all the same if the physician were coding based on time. On the other hand, if a physician codes based on the complexity of the chief complaint, history and physical exam, or medical decision making, this time constraint would be less binding.

These explanations highlight the importance of the interpretation of coding definitions in the Medicare Physician Fee Schedule, which is widely adopted by commercial insurers. Policies that aim to achieve savings by targeting only the price (fees) component of health care spending may be susceptible to coding changes that (at least partially) offset its intended effects. Increased coding intensity can be difficult to separate from inappropriate "upcoding," even after adjusting for health risk scores, because of the flexibilities in code definitions. Given that physicians exert at least some degree of clinical decision making authority, fee-based policies in any underlying fee-for-service payment system may lead to downstream coding (or even volume) effects.

This study is subject to several limitations. First, the MarketScan Medicare Supplemental database contains only beneficiaries who have employer-sponsored Medicare supplemental (Medigap) coverage. Thus, results may not generalize to Medicare beneficiaries without such coverage. The Marketscan population is derived mostly from large employers. Employees of medium and small firms, the unemployed, and other Medicare beneficiaries are not represented. I was also unable to ascertain changes in quality, access, or other aspects of care delivery which may have changed due to the policy.

In addition, the policy's effects in the first year may not generalize to the long-run. For example, the increased volume of PCP visits may lower future spending through preventive care. In addition, the policy's simultaneous increase in office visits fees could have encouraged PCPs to see more complex patients and refer them less to specialists. This mechanism, which is not inconsistent with the results, may affect subsequent care patterns and spending for such patients.

This initial evaluation of Medicare's elimination of consultations in 2010 offers several potential lessons for policymakers. First, redistributing payments between specialties in a budget-neutral manner through fee changes alone may lead to unintended effects that increase overall spending. While the asymmetric fee cut to specialists in the consultations setting did not generate savings among specialists, policies such as the 10-

year proposal by the Medicare Payment Advisory Commission, which imposes broader fee changes across multiple categories of services, may produce different results. Second, the inherent flexibility and subjectivity of code definitions could lead to potentially undesirable coding behavior in response to fee-based policies. There are numerous areas in the fee schedule that feature a gradient of service intensities captured by a set of closely related codes. Third, fee cuts may be overall less effective in generating savings than policies that control total spending such as bundled payments, which remove overt incentives for offsetting fee cuts through coding or volume changes.[27,28]

2.5. References

1. Iglehart JK. Medicare Payment Reform — Proposals for Paying for an SGR Repeal. N Engl J Med 2011; 365:1859-1861
2. Aaron HJ. The Central Question for Health Policy in Deficit Reduction. N Engl J Med 2011; 365:1655-1657.
3. Ginsburg PB, White C. Health Care's Role in Deficit Reduction — Guiding Principles. N Engl J Med 2011; 365:1559-1561
4. Chernew ME, Baicker K, Hsu J. The specter of financial armageddon--health care and federal debt in the United States. N Engl J Med. 2010;362:1166-8.
5. Congressional Budget Office. Changes in Payment to Physicians. 2011 June. (http://www.cbo.gov/doc.cfm?index=12240). Accessed 18 Dec 2011.
6. Centers for Medicare and Medicaid Services. Physician Fee Schedule: PFS Relative Value Files. (https://www.cms.gov/PhysicianFeeSched/PFSRVF/)
7. Hsiao WC, Braun P, Dunn D, Becker ER. Resource-based relative values. An overview. JAMA. 1988 Oct 28;260(16):2347-53.
8. Newhouse JP. Pricing the Priceless: A Health Care Conundrum. Massachusetts Institute of Technology Press: Cambridge, 2002.
9. Leigh JP, Tancredi D, Jerant A, Kravitz RL. Physician wages across specialties: informing the physician reimbursement debate. Arch Intern Med. 2010 Oct 25;170(19):1728-34.
10. Lasser KE, Woolhandler S, Himmelstein DU. Sources of U.S. physician income: the contribution of government payments to the specialist-generalist income gap. J Gen Intern Med. 2008;23(9):1477-1481.

11. Bodenheimer T, Berenson RA, Rudolf P. The primary care-specialty income gap: why it matters. Ann Intern Med. 2007;146(4):301-306.

12. Hackbarth GM. Moving forward from the sustainable growth rate (SGR) system. In: Letter to Senators Max Baucus and Orrin G. Hatch and Representatives Dave Camp, Sander M. Levin, Fred Upton, and Henry A. Waxman. Washington, DC: Medicare Payment Advisory Commission, October 14, 2011.

13. Centers for Medicare and Medicaid Services, Department of Health and Human Services. Medicare Program; Payment Policies Under the Physician Fee Schedule and Other Revisions to Part B for CY 2010. Federal Register 2009 Nov 25;74(226): 61737–62206.

14. Centers for Medicare and Medicaid Services. CMS Manual System, Pub 100-04 Medicare Claims Processing. Transmittal 1875. December 14, 2009.

15. Hansen LG, Chang S. Health Research Data for the Read World: the Thomson Reuters Marketscan Databases. Thomson Reuters White Paper, 2010 Jan. (http://marketscan.thomsonreuters.com/marketscanportal/)

16. Pope GC, Kautter J, Ellis RP, Ash AS, Ayanian JZ, Lezzoni LI, Ingber MJ, Levy JM, Robst J. Risk adjustment of Medicare capitation payments using the CMS-HCC model. Health Care Financ Rev. 2004 Summer;25(4):119-41.

17. Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. The implications of regional variations in Medicare spending. Part 1: the content, quality, and accessibility of care. Ann Intern Med. 2003 Feb 18;138(4):273-87.

18. Welch WP, Miller ME, Welch HG, Fisher ES, Wennberg JE. Geographic variation in expenditures for physicians' services in the United States. N Engl J Med. 1993;328:621-7.

19. Wennberg JE. The Dartmouth Atlas of Health Care 1996. Chicago: American Hospital Publishing; 1996.
20. White H. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 1980;48:817-30.
21. Huber PJ. The behavior of maximum likelihood estimates under non-standard conditions. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967:221-33.
22. Rice TH. The Impact of Changing Medicare Reimbursement Rates on Physician-Induced Demand. Med Care 1983. 21(8):803-15.
23. Nguyen NX, Derrick FW. Physician Behavioral Response to a Medicare Price Reduction. Health Serv Res 1997. 32:3:283-98.
24. Yip WC. Physician response to Medicare fee reductions: changes in the volume of coronary artery bypass graft (CABG) surgeries in the Medicare and private sectors. J Health Econ 1998. 17:675-99.
25. Zuckerman S, Norton SA, Verrilli D. Price controls and Medicare spending: assessing the volume offset assumption. Med Care Res Rev. 1998 Dec;55(4):457-78.
26. Jacobson M, Earle CC, Price M, Newhouse JP. How medicare's payment cuts for cancer chemotherapy drugs changed patterns of treatment. Health Aff (Millwood). 2010 Jul;29(7):1391-9.
27. Chernew M, Goldman D, Axeen S. How much savings can we wring from Medicare? N Engl J Med. 2011 Oct 6;365(14):e29.
28. Kaiser Family Foundation. Medicare: Medicare Spending and Financing. Fact Sheet, August 2010. (http://www.kff.org/medicare/upload/7305-05.pdf)

2.6. Supporting Materials

Contents

Section 2.6.1.

Table 2.3. Change in professional fees between 2009 consultations and 2010 office visits (the effective fee cut)*

Consultation fees in 2009		Office visit fees in 2010		
HCPCS code	**Fee ($)**	**HCPCS code**	**Fee ($)**	**% Change**
99241 (lowest complexity)	48.69	99201 (new) 99211 (estab.)	38.97 19.12	-20 -61
99242	90.89	99202 (new) 99212 (estab.)	67.11 38.97	-26 -57
99243	124.79	99203 (new) 99213 (estab.)	97.77 65.30	-22 -48
99244	184.30	99204 (new) 99214 (estab.)	151.89 97.77	-18 -47
99245 (highest complexity)	226.50	99205 (new) 99215 (estab.)	190.50 132.05	-16 -42

* On the left hand side are consultation codes (99241-99245) and their 2009 fees, which were eliminated in 2010. On the right hand side are 2010 fees for office visits, which physicians were instructed to substitute towards. The % change column describes the change in fees for 2010 office visits compared to 2009 fees for consultations. Codes 99201-99205 are office visits for new patients. Codes 99211-99215 are office visits for established patients. All fees are national Payment Amounts for outpatient professional fees derived from the Medicare Part B Physician Fee Schedule.

Section 2.6.2.

Table 2.4. Change in office visit fees from 2009 to 2010 (the compensatory fee increase)*

HCPCS code	2009 Fee ($)	2010 Fee ($)	% Change
99201 (new)	36.79	38.97	+6
99211 (estab.)	18.75	19.12	+2
99202 (new)	63.48	67.11	+6
99212 (estab.)	37.15	38.97	+5
99203 (new)	91.97	97.77	+6
99213 (estab.)	61.31	65.30	+7
99204 (new)	141.74	151.89	+7
99214 (estab.)	92.33	97.77	+6
99205 (new)	178.89	190.50	+6
99215 (estab.)	124.79	132.05	+6

* This table shows the change in office visit fees from 2009 to 2010. Codes 99201-99205 are office visits for new patients. Codes 99211-99215 are office visits for established patients. All fees are national Payment Amounts for outpatient professional fees derived from the Medicare Part B Physician Fee Schedule.

Section 2.6.3.

Table 2.5. Definitions of clinical complexity for patient encounters

The following definitions are shared by the common procedural definition (CPT) coding system, used by most commercial insurers, and the Healthcare Common Procedure Coding System (HCPCS) used by Medicare. “Key components” and “presenting problem(s)” are identical for consultations, new office visits, and established office visits.

Level of Complexity	**Description**	
1	**Key components**	• A problem focused history; • A problem focused examination; and • Straightforward medical decision making.
	Presenting problem(s)	Self limited or minor
	Time	• Consultation (99241)—15 min • New office visit (99201)—10 min • Established office visit (99211)—5 min
2	**Key components**	• An expanded problem focused history; • An expanded problem focused examination; and • Straightforward medical decision making.
	Presenting problem(s)	Low severity
	Time	• Consultation (99242)—30 min • New office visit (99202)—20 min • Established office visit (99212)—10 min
3	**Key components**	• A detailed history; • A detailed examination; and • Medical decision making of low complexity.
	Presenting problem(s)	Moderate severity
	Time	• Consultation (99243)—40 min • New office visit (99203)—30 min • Established office visit (99213)—15 min
4	**Key components**	• A comprehensive history; • A comprehensive examination; and • Medical decision making of moderate complexity.
	Presenting problem(s)	moderate to high severity
	Time	• Consultation (99244)—60 min • New office visit (99204)—45 min • Established office visit (99214)—25 min
5	**Key components**	• A comprehensive history; • A comprehensive examination; and • Medical decision making of high complexity.
	Presenting problem(s)	moderate to high severity
	Time	• Consultation (99241)—80 min • New office visit (99205)—60 min • Established office visit (99215)—40 min

Section 2.6.4.

Table 2.6. Sensitivity analyses

The following table lists interrupted time series estimates of the 2010 Medicare policy's effect on spending and volume, by type of patient encounter as well as for all encounters. Column headings denote the dependent variable in the model, in units of either dollars per beneficiary per quarter (for spending models—columns 1-4) or volume per 100 beneficiaries per quarter (volume models—columns 5-8). The base model results (row 1) are compared to results from sensitivity analyses (rows 2-6). Sensitivity analyses are alterations from the base model. P values are in parentheses.

	Adjusted change in spending ($ per beneficiary per quarter)				**Adjusted change in volume (per 100 beneficiaries per quarter)**			
	[1] All Encounters	**[2] Consults**	**[3] New office visits**	**[4] Established office visits**	**[5] All Encounters**	**[6] Consults**	**[7] New office visits**	**[8] Established office visits**
Base model	10.20 (p<0.001)	-18.52 (p<0.001)	13.64 (p<0.001)	15.08 (p<0.001)	-1.4 (p=0.18)	-11.4 (p<0.001)	8.7 (p<0.001)	1.3 (p=0.19)
Omit quarter indicators	9.43 (p<0.001)	-18.37 (p<0.001)	13.82 (p<0.001)	13.97 (p<0.001)	-1.4 (p=0.18)	-11.4 (p<0.001)	8.7 (p<0.001)	1.3 (p=0.19)
Omit HRR indicators	10.69 (p<0.001)	-18.46 (p<0.001)	13.69 (p<0.001)	15.46 (p<0.001)	-3.4 (p<0.001)	-11.7 (p<0.001)	8.3 (p<0.001)	0.0 (p=0.85)
Omit linear trend	8.39 (p<0.001)	-19.39 (p<0.001)	13.22 (p<0.001)	14.56 (p<0.001)	3.0 (p=0.002)	-11.5 (p<0.001)	8.6 (p<0.001)	5.9 (p<0.001)
Omit age, sex, and risk score	8.86 (p<0.001)	-18.56 (p<0.001)	13.65 (p<0.001)	13.77 (p<0.001)	-5.2 (p<0.001)	-11.7 (p<0.001)	8.3 (p<0.001)	-1.9 (p=0.04)
Continuous 4-year enrollees	9.93 (p<0.001)	-18.60 (p<0.001)	13.53 (p<0.001)	15.00 (p<0.001)	-2.6 (p<0.001)	-11.6 (p<0.001)	8.4 (p<0.001)	0.5 (p=0.42)

Chapter 3

Competitive bidding in Medicare Advantage:

Effect of benchmark changes on plan bids*

* A version of this chapter has been submitted for publication. Please do not circulate.

3.1. Introduction

Bidding is one of the most important price-setting mechanisms in economics. It is central to auctions, which are commonly used to set prices in the absence of a preexisting market (Hansen, 1988; Vickrey, 1961). A large body of theoretical work has demonstrated that the outcome of markets that use bidding to set prices depends, in part, on the nature of competition in the market (McAfee and McMillan, 1987; Klemperer, 1999). This has been followed by a growing empirical literature on auctions (Athey and Haile, 2006; Hendricks and Porter, 2007). Bidding is especially important in shaping government procurement contracts for the delivery of public services through the private sector (Laffont and Tirole, 1993; Bajari and Tadelis, 2001). In recent years, this context for bidding has assumed an increasingly central role in health care, in particular the United States Medicare Program. The Medicare Part D prescription drug market, the market for durable medical equipment, and the Medicare Advantage, in which commercial insurers contract with Medicare to provide alternative insurance options to standard Medicare Part A and Part B coverage for Medicare beneficiaries, all use bidding as a way to set price.

Increasingly, proposals to control Medicare spending center on competitive bidding as a market based alternative to administratively imposed payment reduction (Antos, 2012; Wilensky, 2012). Competitive bidding is the foundation of a recent legislative proposal by Paul Ryan (R-WI) and Ron Wyden (D-OR)—the Ryan-Wyden Plan—which proposes replacing traditional fee-for-service Medicare with an expanded bidding system (Wyden and Ryan, 2011). Some analysts predict that expanding the role of bidding in Medicare could save $339 billion or 9.5 percent of Medicare spending through 2020 (5.6 percentage points more than projected savings under the Patient Protection and Affordable Care Act) (Feldman, Coulam, and Dowd, 2012). Yet despite the importance of bidding, there is little empirical work to date on plan bidding behavior in MA.

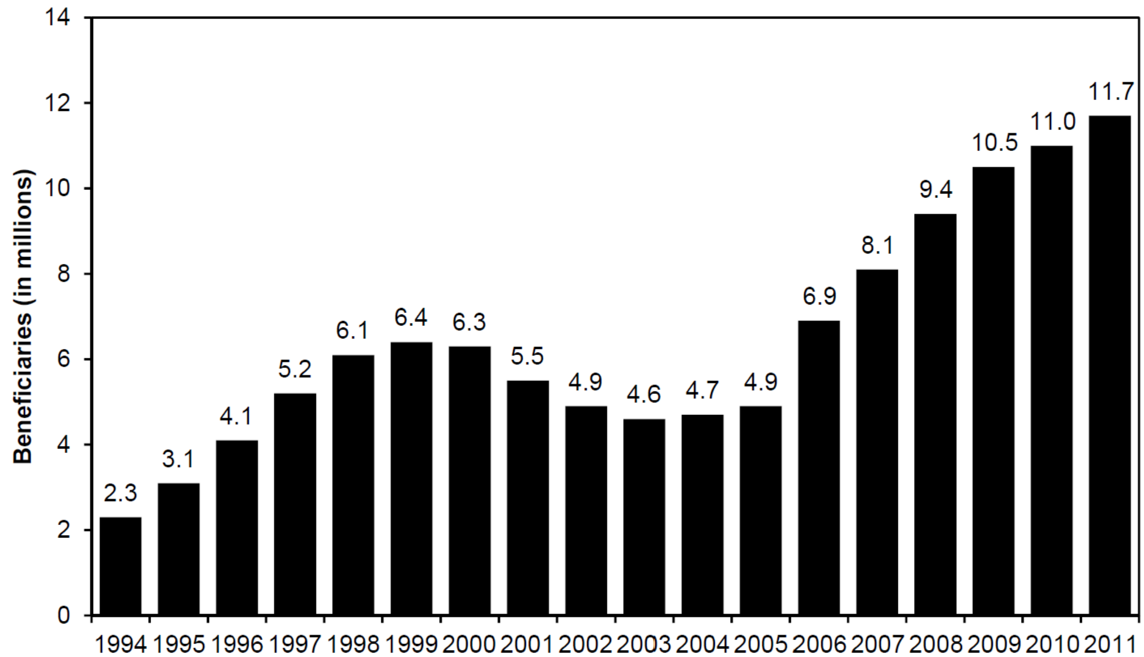
In the MA bidding system, any commercial insurer that would like to offer an MA plan submits a plan-specific bid (an amount covering the expected costs of a standard benefit package for an average risk beneficiary) to the Centers for Medicare and Medicaid Services (CMS). This bid, which must also be accompanied by projected enrollment in the counties covered by the plan, is measured against county-level benchmark rates set by CMS. From the bid and projected enrollment, CMS determines a plan-specific payment. Competition gives plans an incentive to bid low in order to attract consumers. Low bids are salient to beneficiaries through higher “rebates” (additional coverage benefits or reduced cost sharing for Medicare Part B or Part D), upon which beneficiaries’ enrollment decisions are based. As a result, MA plans compete on the basis of price (premiums) and benefits. Given that benchmark rate-setting is the central policy tool for Medicare, this paper examines the following question: How do bids change in response to changes in benchmark payment rates?

Understanding the relationship between county-level benchmark rates and plan bids is also important for impending changes under current law. In 2012, the Patient Protection and Affordable Care Act will begin phasing in a modification of the benchmark rate formula that will likely have substantial effects on plan behavior and MA enrollment. Designed to reduce excessive payments to MA plans, benchmarks will be set as a percentage of average fee-for-service (FFS) Medicare spending in a county, with the percentage determined by how a county's FFS spending compares with that of other counties (MedPAC, 2011). Counties in the highest quartile of FFS spending will face benchmarks equal to 95 percent of average risk-adjusted FFS spending in their region; those in the lowest quartile will face benchmarks equal to 115 percent. Since MA plans are concentrated in the highest spending counties, resulting benchmarks may have large impacts on plan rebates and subsequent MA enrollment.

We study the relationship between benchmark rates and plan bids using plan payment, enrollment, and county FFS spending data from the Centers for Medicare and Medicaid Services (CMS) for the period 2006-2009. The basic strategy is to use a longitudinal model to assess the effect of changes in benchmark rates on changes in plan bids. Since plans locate across multiple counties, I aggregate the data to the "market" level using a weighted average of county level benchmark rates across the counties that each plan enters. I find evidence that bids move with changes in benchmark rates, suggesting that plans do not strictly bid costs. Moreover, I find that plan rebates also move with benchmark rates, but to a lesser degree than bids, suggesting that plans invest part of the extra dollars generated by a higher benchmark on rebates to attract enrollees and retain the rest as profit.

The paper proceeds as follows. Section 2 provides background on the MA program and its competitive bidding system. Section 3 reviews the theory of price-setting in the MA context under perfect and imperfect information. Section 4 describes the data and section 5 presents the empirical strategy. Results are shown in section 6, and section 7 concludes.

Figure 3.1. Growth in Medicare Advantage enrollment, 1994-2011.



Source: Medicare Payment Advisory Commission (2011). A Data Book: Health care spending and the Medicare Program, June 2011. Section 9, p. 147.

3.2. Background

More than 25 percent of Medicare beneficiaries today are enrolled in MA. In the last decade, MA enrollment has more than doubled from 4.5 million in 2003 to 11.4 million in 2010 (MedPAC 2011), its highest level since the inception of the program (Figure 3.1). During this time, MA has commanded increasing policy interest, with growing implications for the trajectory of Medicare spending as well as beneficiary access and quality of care (Gold, 2012; Guram and Moffit, 2012). Beginning in 2006 MA plan payments were determined via a competitive bidding system, which sought to leverage market forces to encourage more efficient and higher quality plan options (McGuire, Newhouse, and Sinaiko, 2011).

The MA program, formerly known as Medicare Part C and Medicare+Choice, provides Medicare beneficiaries the choice of health insurance plans that provide Medicare coverage offered by commercial insurers in lieu of traditional FFS Medicare. Private insurers in MA accept prospective payments that vary at the county level and agree to provide coverage for Medicare Part A and Part B services. Plans may also supply prescription drug, or Medicare Part D, coverage. Commercial plans that enter MA compete in a regulated market. The original motivation of the 1982 Tax Equity and Fiscal Responsibility Act (TEFRA), which ushered in managed care plans in Medicare, was to encourage private plans to improve efficiency and lower Medicare spending. Another important tenet was giving beneficiaries choice to choose the best plan that fits their needs. Plans have flexibility to contract with providers, use managed care techniques, and design beneficiary incentives.

A number of studies have examined the relationship between plan payments and MA enrollment (see, for example, Atherly et al., 2004; Cawley et al., 2005; Dowd et al., 2003; Gowrisankaran and Town, 2006; Maruyama, 2011; Town and Liu, 2003). They use different datasets across varying time horizons and different estimation techniques, and find largely heterogeneous elasticities of MA enrollment with respect to plan payments. In addition, several studies find that payment rates affect the number of HMO plans that enter the market (Cawley et al., 2002; Town and Liu, 2003; Pizer and Frakt, 2002). However, most of this literature predates the introduction of competitive bidding in 2006. Since there, there is little work on the relationship between benchmarks and enrollment

(Chernew et al., 2008), and virtually no work on the effect of benchmarks on plan bidding or plan rebates, largely owing to the lack of available data.

3.2.1. Competitive bidding in Medicare Advantage

The Medicare bidding system can be summarized in the following four steps (Figure 3.2). Let j denote plans and k denote counties.

(1) CMS sets a benchmark payment rate for each county, $bench_k$. The benchmark is a function of lagged fee-for-service spending and a series of update rules, such as a minimal update factor and county floor payment rates.

(2) Each plan j submits a single sealed bid, bid_j, and projected enrollment in counties the plan will serve.

(3) After reviewing a plan's bid and projected enrollment, CMS assigns plans a plan-specific payment benchmark, $Bench_j$. This payment benchmark is an average of county benchmark rates weighted by projected plan enrollment across counties.

Together, the plan payment benchmark ($Bench_j$) and plan bid (bid_j) determine the rebate or premium that a plan will offer or charge enrollees, respectively. If the plan's bid exceeds its benchmark, the plan payment (pay_j) equals its payment benchmark and the plan must collect the difference through enrollee premiums ($premium_j$).

Case 1: $bid_j > Bench_j$

$$pay_j = Bench_j \tag{1}$$

$$premium_j = bid_j - Bench_j \tag{2}$$

On the other hand, if the plan's bid is less than its payment benchmark, the plan receives its bid. In addition, it receives 75 percent of the difference between its payment benchmark and bid as a rebate ($rebate_j$), which it must return to enrollees. There are no premiums charged to enrollees. The rebate may comprise of extra benefits (such as dental and vision care) or lower Part B or Part D premiums. Essentially the rebate is a negative premium. The remaining 25 percent of the difference is retained by Medicare as an effective tax (Figure 2).

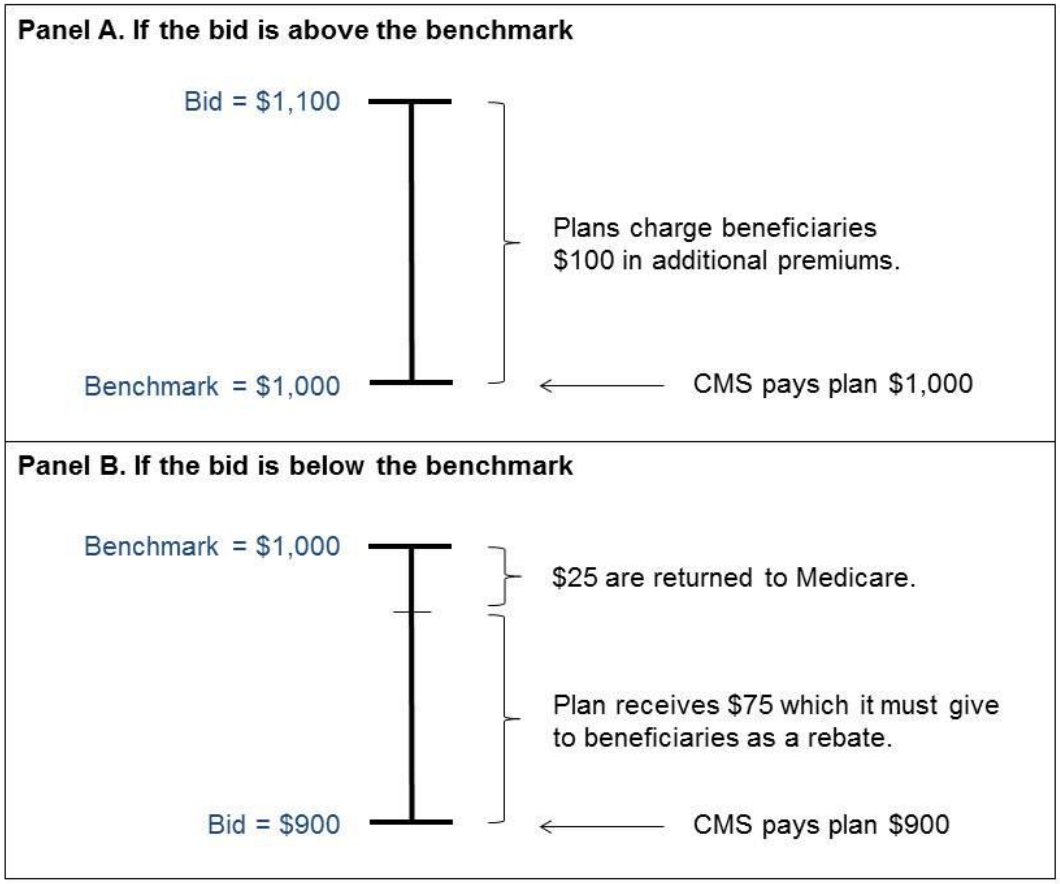
Case 2: $bid_j < Bench_j$

$$pay_j = bid_j + rebate_j = bid_j + 0.75(Bench_j - bid_j) \tag{3}$$

$$premium_j = 0 \tag{4}$$

(4) Given a set of plan options, which differ in premiums and rebates, Medicare beneficiaries choose to enroll in a plan (or in traditional Medicare). The beneficiary's enrollment can be modeled as a discrete choice function and is outside the scope of the current paper.

Figure 3.2. The relationships between the benchmark, bid, and payment.*



* Medicare Advantage plans submit a price (bid) that they are willing to accept from Medicare for insuring a beneficiary. Plan payments are determined by comparing the bid to a "benchmark" payment rate set by CMS based on the counties the plan serves. If a plan bids above the benchmark it faces, the plan receives only the benchmark amount and must collect the difference by charging a premium to enrollees. If a plan bids below the benchmark, it receives its bid plus 75 percent of the difference between the bid and benchmark—a "rebate"—which the plan must return to beneficiaries in the form of lower premiums or additional benefits. The lower the bid, the higher the rebate that plans may use to attract beneficiaries. For more details of the bidding system, see: Medicare Payment Advisory Commission. *Medicare Advantage Program Payment System: Payment Basics*. October 2008, Figure 2 (p. 3).

3.3. Theory

Unlike a typical first-price sealed bid auction, bidding in MA does not have the winner-take-all feature. Instead, the MA bidding system is simply a way to set plan payments and premiums. There are many ways to design a bidding system, and different models of competition generate different predictions. I highlight a benchmark case below.

3.3.1. Perfect competition (Bertrand pricing)

In the standard case of perfect competition, assume MA plans maximize a simple profit function characterized by markups (price above cost) and enrollment (quantity demanded).

$$\pi_j = \Sigma_k(p_j\text{-}c_{jk})n_{jk} = \Sigma_k(p_j\text{-}c_{jk})s_{jk}m_k \qquad (5)$$

where p_j equals bid if the bid is less than the benchmark, c_{jk} denotes plan costs by county (which may be endogenous), and n_{jk} is plan enrollment by county. Note that p_j does not vary by county as plans receive a single price. Enrollment can be expressed as the product of the plan's market share in a county, s_{jk}, and the number of Medicare beneficiaries in a county, m_k.

Under perfect competition, products are perfect substitutes and sellers engage in sharp price competition. Each sets its price simultaneously and non-cooperatively, and the lowest price taking the whole market. In equilibrium, competition forces bids down to plan costs ($bid_j = c_j$), and plan profit is zero ($\pi_j = 0$). Any exogenous change in the benchmark can be modeled as a shift in demand. For example, in the case where the bid exceeds the benchmark, a 1 dollar increase in benchmark will reduce the premium charged to beneficiaries by one dollar. For bids below the benchmark, a 1 dollar increase in benchmark will increase the rebate (e.g. negative premium) by 75 cents. Because the majority of plans (over 90%) bid below the benchmark, this is the dominant case. In any model of perfect competition, the impact of a change in benchmark on the bid will depend on the slope of the supply curve. There is some evidence that suggests the supply curve in the health insurance industry is flat.

This simple model of perfect competition, however, may not be realistic. For example, it implies that anytime such competition exists, a plan pricing marginally lower than competitors will capture all market demand. The model also assumes no friction (search or switching costs) on the demand side, implying that beneficiaries can freely switch between plans. In addition, it assumes plan costs are a function of beneficiary characteristics. If health providers react to a benchmark increase by demanding higher reimbursement, the supply curve would be upwards sloping. In other words, the predictions are specific to a world with no provider market power.

3.3.2. Imperfect competition (Nash pricing)

Many models deviate from perfect competition. The theory of differentiated products oligopoly offers one view, where products are not perfect substitutes and their differentiability confers market power to their suppliers. The classic duopoly theory is attributed to Dixit (1979). In this classic theory with constant marginal costs, Bertrand competition yields lower prices and higher welfare than Cournot competition. If goods are substitutes, supplier profits are lower under Bertrand competition than under Cournot. If goods are complements, Bertrand profits are higher. Thus, the firm's dominant strategy is to choose quantity as its strategic variable when the goods are substitutes, and to choose price when the goods are complements (Singh and Vives 1984).

In the MA context, one can consider what the optimal markups should be in a standard industrial organizations framework. Given differentiated products, plans maximize their profits taking other insurers' prices (bids) as given. The solution depends on the elasticity of demand, nature of competition, and the share of the difference between the bid and the benchmark rate returned to Medicare (25 percent under current law). Benchmark rates affect bids because they shift the demand curve. For example, if a plan's equilibrium net benefit package to enrollees is x*, and the benchmark increases, the optimal plan bid to achieve x* may also rise. This exemplifies the standard result from models of imperfect competition. In the short run, a shift in demand will increase prices even if the marginal cost curve is perfectly elastic.

Another potential source of imperfect competition arises in the provider market. Specifically, provider market power may affect plan bids. Unlike Medicare, which sets its prices, commercial insurers negotiate prices with providers. Provider market power can generate wide variations in commercial insurer prices even within local markets; a good example is Eastern Massachusetts (Office of the Massachusetts Attorney General, 2010). CMS benchmark rates may affect MA plans' reservation provider fees. If providers know that plans will receive higher benchmarks, they may negotiate for higher fees to extract a part of that rent. On the other hand, benchmark rates also affect enrollment, which is directly related to insurer market power. Thus, greater enrollment through increased net benefits (following from increased benchmarks) could lower provider fees.

3.4. Data

We used public CMS databases containing MA plan payment data at both the county and plan level from 2006-2009. I also used county-level published benchmark rates and actual FFS costs per beneficiary. Benchmarks in year *t* are derived by trending average county FFS spending from year *t-8* to year *t-3* forward using the growth rate in that 5-year period. In addition, I used public CMS databases on plan location and enrollment by county in each year. I merged these public datasets to create market-level data, as described below.

The county-level data contains average plan payments, rebates, and county risk scores. Risk scores are the Centers for Medicare and Medicaid Services Hierarchical Condition Category (CMS-HCC) risk scores. The plan-level data contains plan payments, rebates, and plan risk scores. The plan's payment rate is the plan's bid standardized to a beneficiary of 1.0 risk. When plans submit a bid to CMS, they submit a bid based on the expected risk profile of their enrollees as expected enrollment in the counties they serve. CMS subsequently converts this bid to a standardized bid, which is what I use in the analyses. Published county benchmark payment rates are drawn from the MA Ratebook. Specifically, they are the overall "risk" rates for each county, standardized to a beneficiary of 1.0 risk. County level realized FFS costs up to 2009 are also available online through CMS. Like the prior variables, these costs are standardized to the 1.0 risk enrollee. Enrollment is the total number of beneficiaries enrolled in a MA plan.

We restrict analysis to local plans, which use the same county benchmarks. Specifically, I conduct these analyses at three levels of plan aggregation: health maintenance organization (HMOs) only, HMOs and local preferred provider organization plans (LPPOs), and HMOs, LPPOs, and private fee-for-service plans (PFFS). Employer plans and special needs plans are excluded, as are regional PPO plans that use different regional benchmarks.

3.5. Empirical strategy

3.5.1. Level of analysis

Both county-level and plan-level analyses of the relationship between benchmark rates and bids are inherently limited. Because MA plans are required to submit a single bid covering all counties they serve, relating the county benchmark to average bid in any given county ignores the effect of benchmarks in other (often adjacent) counties which may have affected the given county's bid. Alternatively, a plan-level longitudinal model requires plans to be a stable unit of observation across time. Yet insurers frequently consolidate or split their plans from one year to the next. For example, a "silver" plan's enrollees in one year may be subsumed under a "gold" plan in the next year, retaining the gold plan's identification number. Alternative scenarios are also common. The two plans offered by the same insurer may merge and take on a new identification number altogether. Subsequently, this consolidated plan may split in the next year. Such plan dynamics complicate longitudinal fixed effects models with plans as the unit of the observation.

To combat these limitations, I constructed, for each county, an effective “market” that takes into account the benchmarks in other counties that plans observe when making their bids in the given county. Each unique market is centered on a unique county, with market boundaries defined by the set of other counties served at any point in the study period by plans in the central counties. Like counties, they are a stable unit of analysis over time, allowing for longitudinal models. Specifically, I constructed markets from plan- and county-level data in three steps.

First, for each county, I identified all other counties served by plans in the target county. In each year I examine the distribution of enrollment across counties for all plans serving the target county. Second, I assigned the county benchmark payment rate in all non-target counties to the target county, weighted by the share of each of the target county plan’s enrollment in the non-target counties. I used each plan’s average share across the study period so that the weights were stable. This procedure was repeated for other variables. The resulting units of observation, which I call “markets,” are stable over time, allowing for longitudinal models. There is the same number of markets as counties. Each observation in the final dataset is a unique market-year, and I will refer to them as markets in the rest of the paper.

3.5.2. Empirical model

We estimate the effect of changes in benchmark rates on changes in plan bids using the following reduced form county-level model:

$$bid_{kt} = \beta_1 bench_{kt} + \beta_2 FFS\_TM_{kt} + \beta_3 risk\_TM_{kt} + \beta_4 risk\_MA_{kt} + \beta_5 insurers_{kt}$$
$$\beta_6 population_{kt} + \lambda_k + \tau_t + \varepsilon_{kt} \qquad (6)$$

where bid_{kt} is the average bid in market k in year t, $bench_{kt}$ is the market-level published benchmark rate in year t, and FFS_TM_{kt} is the actual realized (contemporaneous) FFS costs of the county's traditional Medicare (TM) population. I also control for health status through the CMS-HCC risk score: $risk\_TM_{kt}$ is the average risk score of the county's TM population, and $risk\_MA_{kt}$ is the average risk score of the county's MA population. In addition, I include the number of insurers in a county and the total population of Medicare eligible persons in a county. County fixed effects are represented by λ_k and year fixed effects by τ_t. County fixed effects render the identification of the effect of benchmark rates on bids through within-county changes. This specification is preferable to cross-sectional or pooled OLS to the extent that benchmark rates are correlated with unobservable county-level characteristics that also affect plan bids. Year fixed effects control for any underlying trends in bids across all counties. In the base model, the error term ε_{kt} is assumed to be uncorrelated with the independent variables. I address the potential endogeneity of the error term below. Observations are weighted by a fixed average of the total Medicare population in a county across 2006-2009. Given that markets overlap geographically, standard errors are clustered by state rather than by county.

To test the robustness of this base model, I begin with a number of basic sensitivity analyses. To assess the impact of selection effects, I repeat the model while omitting TM risk or plan risk. I then repeat analyses using a linear trend in place of year dummies. Next, I repeat the model omitting fixed county weights and restricting attention to big counties. I also include an interaction between benchmark and the number of insurers. Finally, to control for provider market power, I include a Herfindahl–Hirschman Index (HHI) based on the number of hospital beds in a county in one specification, and present an additional specification with the HHI and the interaction between HHI and benchmark.

Since I exploit variation in benchmark changes, which are in dollars in the base model, I also explore a model in which I logarithmically transform bids and benchmarks. Variation in dollars may arise even when the percent change in the benchmarks are the same—for example when multiple counties with different base benchmarks receive the same minimum update of 2 percent or national traditional Medicare growth rate in a given year. The log-transformed model is a generalized linear model (GLM) which allows a more flexible functional form and removes the need to smear the estimates, which is required in the OLS case. I transform the predictions of the GLM into dollars and find that they are very consistent with the base model.

3.5.3. Potential endogeneity

Despite controlling for county FFS costs, the error term in equation (6) may be correlated with the benchmark. There may be unobservable time-varying county characteristics that are correlated with the benchmark and bid. In particular, I are concerned about omitted plan costs in the error term, ε_{kt}. Since I run a longitudinal model with county fixed effects, I can express the terms in equation (6) as changes, where FFS_MA_{kt} is the average plan costs in a county.

$$\Delta bid_{kt} = \beta_1 \Delta bench_{kt} + \ldots + \Delta\varepsilon_{kt} \tag{7}$$

$$\Delta bid_{kt} = \beta_1 \Delta bench_{kt} + \ldots + \Delta(FFS\_MA_{kt} + \mu_{kt}) \tag{8}$$

Since the benchmark in year t is calculated using a linear projection of the county's FFS spending in a 5-year period from $t\text{-}8$ to $t\text{-}3$, I can express $\Delta Bench_{kt}$ as the following

$$\Delta bench_{kt} = bench_{kt} - bench_{k,t-1} \tag{9}$$

$$\Delta bench_{kt} = f(FFS\_TM_{k,t-3}, \ldots, FFS\_TM_{k,t-8}) - f(FFS\_TM_{k,t-4}, \ldots, FFS\_TM_{k,t-9}) \tag{10}$$

Thus, it is clear from equation (10) that there are 4 overlapping years, $FFS\_TM_{k,t-4}$ through $FFS\_TM_{k,t-8}$, which determine the benchmarks in both year t and $t\text{-}1$. Thus, the concern is whether the omitted variable—the change in plan costs from year t to $t\text{-}1$—is correlated with a function of the FFS costs in year $t\text{-}3$ and $t\text{-}9$.

$$\Delta bench_{kt} = g(FFS\_TM_{k,t-3}, FFS\_TM_{k,t-9}) \tag{11}$$

We address this potential endogeneity in several ways, including using instrumental variables and assessing the sensitivity to omitting FFS costs (which should increase the bias if omitted costs I a large concern). I also estimate models on 2006-2007 and 2008-2009 separately, because payment updates in those years ("rebasing years") were dominated by legislative changes to payment floors that are plausibly more exogenous.

3.5.4. Instrumental variables

In order to address this bias, I use simulated benchmark rates to identify the effect of changes in benchmarks on bids. I would expect OLS estimates of β to be biased upwards if bids tend to increase more in areas where benchmarks increase more. On the other hand, β would be biased towards zero if changes in benchmark changes have an inverse relationship with changes in bids. With imperfect competition, I might expect bids to increase more in areas greater benchmark increases, which offer greater rents to be captured by the plan or providers.

We construct simulated benchmarks by isolating the plausibly exogenous portion of benchmark updates in each year, and using that portion to update the (simulated) benchmark from the previous year. From 2006-2009, benchmarks were updated by the maximum value of 4 different paths: (1) a minimum update of 2 percent, (2) the national FFS Medicare growth rate, (3) the county FFS growth rate, or (4) an urban or rural floor update. Since the county's own FFS growth rate (path 3) is potentially endogenous, I abolish this criterion and replace it with the FFS growth rate of the entire *state* that a given county belongs to, calculated using a weighted average of county growth rates in all counties in that state *except* the given county. One caveat, however, is that a state-level growth rate update contains, for any given target county, spending in non-target counties, which are often in the same state. Therefore, a simulated benchmark that undergoes this "state-not-self" update will likely be correlated with the error term in the second stage model. Empirically, the state-not-self path is invoked rarely (133 counties were updated via this path in 2007, and 159 counties in 2009). Overall, to the extent that simulated benchmarks are correlated with observed benchmarks but uncorrelated with the error, the IV estimate would represent an improvement over the base OLS estimate.

While the effect of simulated benchmark changes on bids may be less related to the error term, simulated benchmarks may still be a weak instrument if simulated updates are uncorrelated with actual benchmark updates. I test the strength of the instrument using a standard F-test, which strongly rejects the hypothesis that simulated benchmarks are uncorrelated with the actual benchmarks.

3.6. Results

3.6.1. Descriptive analysis

From 2006-2009, there were a total of 11,500 unique Medicare Advantage plans. After excluding regional PPOs, special needs plans, and employer-sponsored plans, Figure 3.3 shows the number of HMO, LPPO, and PFFS plans in each year. While less managed plans have grown in recent years, HMO plans remain the dominant plan type. I focus on HMO plans in the main analyses, which are also the dominant plan type in the most populated regions on the country. While 2010 is included in this descriptive analysis, I exclude it from the econometric analysis because actual realized FFS costs are not yet available for 2010. Figure 3.4 shows the population-weighted average market benchmark rates and average HMO plan bids. In econometric analysis, the identification comes from comparing how bids change in markets with large benchmark changes to how they change in markets with small benchmark changes.

Figure 3.3. Number of Medicare Advantage plans, 2006-2010

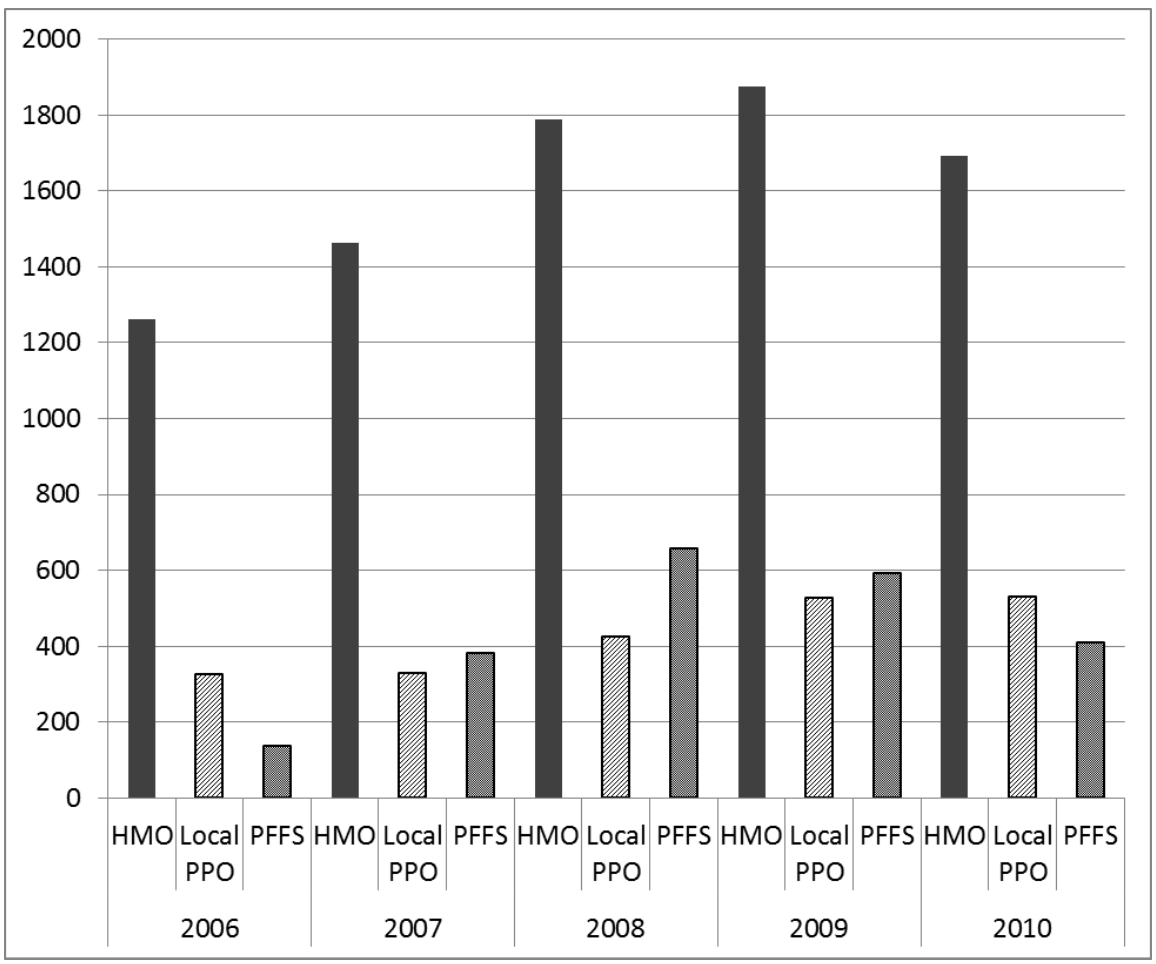
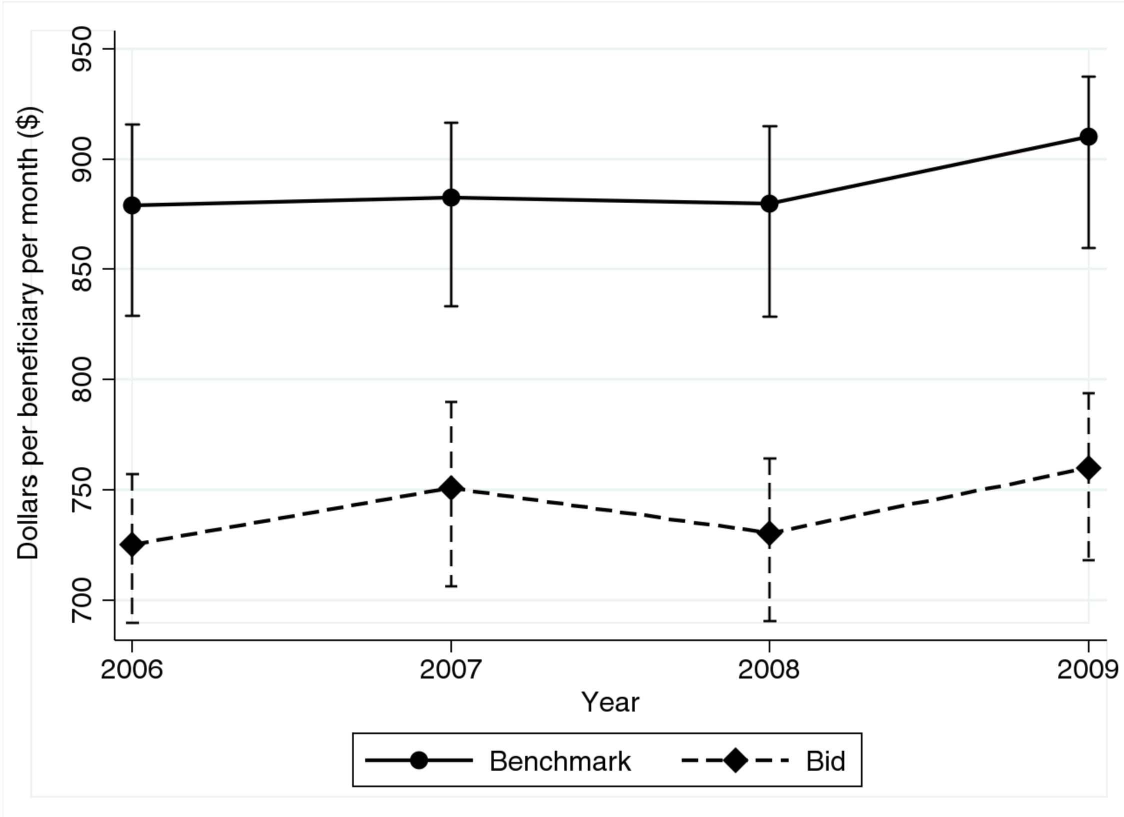


Figure 3.4. Average market benchmarks and bids, 2006-2009



Population-weighted average published benchmark rates and plan bids at the market level standardized to a beneficiary of 1.0 risk, with interquartile range (2012 U.S. dollars).

3.6.2. Econometric results

We find a positive and robust effect of benchmark changes on changes in plan bids. Table 3.1 presents the main OLS estimates and sensitivity analyses for HMO plan markets. I find in the base specification (column 1) that a one dollar increase in the benchmark leads to a $0.50 increase in bids. This result is largely robust to sensitivity analyses. Omitting plan enrollee risk or FFS beneficiary risk did not appreciably change the main estimate, which suggests that differential selection did not play a major role (columns 2-3). Models using a linear yearly trend, omitting the fixed population weights, and restricting attention to large markets also produced estimates near $0.50 (columns 4-6).

We find a fairly consistently negative effect of competition (number of insurers) on bids across different specifications. I would expect this finding given that plans have an incentive to bid low in order to offer a higher rebate (extra benefits) to attract enrollees. On average, each market had 12.1 insurers in 2009, up from 7.9 in 2006, suggesting overall competition increased over these years. Analogously, the average number of insurers in a county (disaggregated from the market-level analysis) grew from 5.7 to 10.2 over this period, giving a sense of the overlap of counties within markets. However, this independent effect largely goes away when I include an interaction between the number of insurers and benchmark (column 7). The negative sign on the interaction term, while small in magnitude and marginally significant, suggests that bids are less responsive to benchmarks in areas with more insurers. This is consistent with the interpretation of the positive relationship between bids and benchmarks being a sign of imperfect competition.

We also find that the relationship between risk in the MA plans and benchmarks is consistently negative, though generally not statistically significant. The negative coefficient is consistent with the notion that the cost advantage of MA plans rises with beneficiary risk. Thus, MA plans would be paid more to care for higher risk enrollees than they need. Since plan bids are for a standard enrollee, and plans are paid more if they attract sicker enrollees, plans would be able to bid lower if they anticipated attracting a sicker population.

Table 3.1. Main estimates and sensitivity analyses (market-level model).

VARIABLES	(1) Base	(2) No county risk	(3) No plan risk	(4) Linear year	(5) Unweighted	(6) Large markets	(7) Interaction	(8) HHI	(9) HHI*bench
Benchmark	0.495***	0.494***	0.486***	0.533***	0.422***	0.523***	0.593***	0.464***	0.526***
	(0.0546)	(0.0549)	(0.0541)	(0.0482)	(0.0633)	(0.0487)	(0.0707)	(0.0509)	(0.0676)
Risk (FFS)	-10.25		-12.20	-29.88*	-5.712	-9.850	-10.07	-8.776	-8.660
	(16.43)		(17.15)	(15.14)	(13.74)	(17.00)	(16.28)	(16.24)	(16.31)
Risk (Plan)	-28.91	-31.44		-49.83**	-21.24	-26.71	-32.38	-28.75	-27.85
	(22.15)	(23.03)		(20.80)	(19.95)	(23.16)	(22.17)	(22.12)	(21.98)
FFS cost	0.0585	0.0557	0.0570	0.0596	0.130**	0.0384	0.0683	0.0611	0.0591
	(0.0478)	(0.0470)	(0.0475)	(0.0450)	(0.0618)	(0.0447)	(0.0490)	(0.0456)	(0.0449)
Pop (1000s)	-0.00835	-0.00670	-0.00940	-0.000994	0.0394	-0.0125	-0.00327	-0.0265	-0.0291
	(0.0362)	(0.0368)	(0.0359)	(0.0366)	(0.0352)	(0.0365)	(0.0353)	(0.0393)	(0.0391)
Insurers	-1.173***	-1.181***	-1.208***	-1.436***	-1.812***	-1.082***	2.509	-0.983***	-1.013***
	(0.285)	(0.289)	(0.285)	(0.289)	(0.347)	(0.283)	(2.122)	(0.296)	(0.318)
Insurers*Bench							-0.00423*		
							(0.00248)		
HHI (beds)								-0.00825***	0.0206
								(0.00272)	(0.0232)
HHI*Bench									-3.26e-05
									(2.48e-05)
Market FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y		Y	Y	Y	Y	Y
Year (linear)				4.038**					
				(1.583)					
Constant	307.3***	302.3***	290.3***	-7,779**	306.8***	294.3***	217.9***	346.0***	292.5***
	(45.96)	(45.19)	(47.00)	(3,168)	(46.19)	(44.95)	(63.19)	(44.80)	(67.25)
Observations	6,036	6,036	6,036	6,036	6,036	4,280	6,036	6,036	6,036
R-squared	0.360	0.359	0.358	0.332	0.317	0.373	0.363	0.367	0.368
Markets	2,013	2,013	2,013	2,013	2,013	1,447	2,013	2,013	2,013

Robust standard errors in parentheses clustered by state. *** p<0.01, ** p<0.05, * p<0.1

We find that bids are inversely related to provider market HHI (column 8). The HHI is a measure of provider concentration in a market, with higher HHI representing greater concentration (or less competition). Having more providers in a market (lower HHI) may lead to higher costs because providers compete on expensive services (i.e. the medical arms race). However, the HHI is not correlated to the relationship between benchmark and bid (column 9). The effect of benchmark changes on bids is largely robust to the inclusion of provider market HHI and the HHI-benchmark interaction.

The coefficient on benchmark in the logarithmically transformed GLM model indicated that a 1 percent increase in benchmark leads to a 0.62 percent (S.E. 0.07) increase in bids (not shown). Transforming this estimate into dollars gives a prediction of $0.52, consistent with the base model and sensitivity analyses. As a further sensitivity analysis, I estimated a similar set of longitudinal models at the *plan* level on the subset of continuously existing plans (stable plans with the same identification number) over the 4 years. Results are largely consistent with market-level models (Table 3.2).

In the supporting materials section, I present analogous main estimates and sensitivity analyses when LPPO and PFFS plans are included in the market. In general, I find that the effect of benchmark changes on bids is attenuated ($0.47 with LPPOs, $0.34 with LPPOs and PFFS) when less managed plans are included in the analysis. These plans, particularly PFFS plans, are largely located in small or rural counties.

Table 3.2. Additional sensitivity analyses (plan-level model).*

	(1) HMO only	(2) HMO and PPO	(3) HMO, PPO, and PFFS
Benchmark ($)	0.605***	0.612***	0.620***
	(0.131)	(0.125)	(0.118)
Plan enrollee CMS-HCC risk score	12.75	11.75	38.24**
	(19.23)	(18.18)	(18.88)
Fee-for-service Medicare cost ($)	-0.0328	-0.0271	-0.00640
	(0.0291)	(0.0281)	(0.0274)
Number of enrollees	0.223	0.170	0.158
	(0.290)	(0.323)	(0.287)
Number of competing insurers	1.033	0.964	1.502
	(1.126)	(1.072)	(0.989)
Benchmark * competitors	-0.00109	-0.00103	-0.00163
	(0.00131)	(0.00125)	(0.00116)
Plan FE	Y	Y	Y
Year FE	Y	Y	Y
Constant	199.0	196.1	145.0
	(130.5)	(125.1)	(115.9)
Observations	2,436	3,140	3,432
R-squared	0.467	0.453	0.471
Number of plans	609	785	858

Robust standard errors in parentheses, clustered by plan. *** p<0.01, ** p<0.05, * p<0.1.

* These sensitivity analyses use longitudinal plan-level models. Thus, each model includes plan indicators (plan fixed effects). Only stable plans (those existing in each year from 2006-2009) were included. Each model was also weighted by plan enrollment size. Column (1) consists of HMO plans only. Column (2) adds local PPO plans to the sample. Column (3) includes all plans (HMO, local PPO, and PFFS). Standard errors are clustered at the plan level.

3.6.3. Sensitivity analyses related to endogeneity

Our major concern is that the change in benchmark is related to unobserved changes in costs. This is unlikely because I condition on actual FFS costs and the benchmark is related to lagged costs. Nevertheless, the correlation between benchmark levels and FFS cost levels is 0.68. However, the correlation of *changes* in benchmarks and *changes* in FFS costs is 0.36. This is similar to the correlation of the residuals of benchmarks and FFS costs (0.37). I conduct several sensitivity analyses to explore the potential bias due to such endogeneity.

Table 3.3 shows the main sensitivity tests related to endogeneity, with the base model replicated in column 1. First, I estimate the base model omitting actual FFS costs (column 2). If benchmark changes were strongly related to costs, I would expect that omission of FFS costs to create a stronger bias and affect the coefficient. Yet, when I omit FFS costs, the benchmark effect on bids is only slightly larger ($0.53).

Second, in 2006-2007 and 2008-2009, the updating of benchmarks underwent a "rebasing," which is required by law to occur at least once every three years. The rebasing of benchmarks differentially increases lower benchmarks by more than higher ones, but is less likely tied to the annual change in costs. When I repeat the base model on just the 2006-2007 data, the effect of changes in benchmarks on bids is $0.54 (S.E. = $0.10) (column 3). Similarly, restricting the model to the 2008-09 data produces an effect of $0.56 (S.E. = $0.06) (column 4). The 2008-2009 period is particularly relevant because benchmark updates in 2060 of the approximately 3200 counties were determined by changes in floor payments, which are most likely exogenous to market level changes in cost.

Finally, the IV estimate closely approximates the base model estimate, demonstrating that a dollar increase in benchmarks leads to about a $0.48 increase in bids and suggesting that the OLS estimate may be biased somewhat upwards (column 5).

Table 3.3. Sensitivity analyses concerning endogeneity (market-level model).*

Variables	**(1)** **Base**	**(2)** **No FFS**	**(3)** **2006-07**	**(4)** **2008-09**	**(5)** **IV**
Benchmark	0.495***	0.531***	0.536***	0.560***	0.481***
	(0.0546)	(0.0515)	(0.0986)	(0.0629)	(0.0782)
Risk (FFS)	-10.25	-8.732	30.63	-188.1**	-10.17
	(16.43)	(16.03)	(18.33)	(89.20)	(10.09)
Risk (Plan)	-28.91	-27.86	-121.9***	13.73	-27.95
	(22.15)	(22.05)	(36.69)	(25.07)	(19.99)
FFS cost	0.0585		-0.0156	-0.000977	0.0641
	(0.0478)		(0.0877)	(0.0530)	(0.0516)
Pop (1000s)	-0.00835	-0.00404	-0.0732	0.0700	-0.00758
	(0.0362)	(0.0377)	(0.0757)	(0.0434)	(0.0121)
Insurers	-1.173***	-1.120***	-0.728*	-2.427***	-1.172***
	(0.285)	(0.268)	(0.425)	(0.435)	(0.183)
Market FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Constant	307.3***	316.3***	382.7***	457.7***	
	(45.96)	(43.32)	(76.99)	(114.4)	
Observations	6,036	6,036	2,808	3,228	5,647
R-squared	0.360	0.357	0.355	0.515	0.360
Markets	2,013	2,013	1,659	1,809	1,624

Robust standard errors in parentheses clustered by state. *** p<0.01, ** p<0.05, * p<0.1

3.6.4. Effect on rebates

Since bids are directly related to rebates, the effect of benchmark changes on rebates follows mechanically from its effect on bids. For every dollar increase in benchmarks, I would expect bids and rebates to move in the same direction as long as the bid increases by less than a dollar. In other words, a dollar increase in benchmarks must translate into a dollar increase from the sum of changes in bids, rebates, and the amount of the 25 percent tax. I would expect the following identity to hold from the model's estimates. $0.75(1 - \hat{\beta}) = \Delta rebate$. The results predict that a \$1 increase in the benchmark would lead to a \$0.38 increase in rebates. As a confirmatory analysis, the base model with rebates on the left-hand side produced an estimate of \$0.34 (S.E. \$0.04).

3.7. Conclusion

In this paper, I examine the competitive bidding system in MA, a relatively unstudied component of the MA program. I assessed the effect of benchmark changes on plan bids and rebates, showing that a \$1 increase in benchmark leads to a \$0.50 increase in bids among HMO plans. Given that the supply curve is likely very elastic, models of perfect competition predict that bids should be insensitive to changes in the benchmark .However, I demonstrate that bids consistently move with benchmarks even after controlling for actual FFS costs and both plan (MA) and FFS beneficiary risk. Sensitivity analyses and an instrumental variable approach using simulated benchmark updates consistently produced similar estimates. To my knowledge, this is the first empirical study of the competitive bidding system in Medicare.

Our results are consistent with several possible explanations. First, they are consistent with economic models of imperfect competition, in which insurers exercise their market power and use higher bids to boost profits at the expense of rebates to beneficiaries. Second, these findings are also consistent with providers exercising market power in their negotiations with insurers. Specifically, providers may observe (or anticipate) CMS increases in the benchmark rates and capture some of the increase through negotiating for higher fees from the commercial plans. I cannot precisely distinguish between these explanations.

This study has several limitations. Importantly, I do not observe actual plan costs. While I control for contemporaneous FFS costs in the model, unobserved plan costs could be different. For example, the risk profile of MA beneficiaries may be different from that of FFS Medicare beneficiaries, and MA plans may indirectly select for healthier beneficiaries through benefit design, advertising, or other means (Brown et al., 2011; Newhouse, 1996). I attempt to address this concern by controlling for FFS beneficiary and MA plan enrollee risk.

Our findings have potential implications for proposals to expand the role of competitive bidding in Medicare, such as the Ryan-Wyden plan and Domenici-Rivlin proposal. These results suggest that competitive bidding markets in MA are not perfectly competitive and may not drive plan bids down to plan costs. Thus, they may temper the projected savings of replacing Medicare with a bidding system (Feldman, Coulam, and Dowd, 2012). Other models of bidding, including those in which the benchmark is not known in advance, may yield different results. As the debate over Medicare reform escalates, the design of market-based bidding mechanisms should be done with these cautions in mind.

3.8. References

Antos, J.R. (2012). The Wyden–Ryan Proposal — A Foundation for Realistic Medicare Reform. New England Journal of Medicine, 366, 879-881

Athey, S. and Haile, P.A. (2006). Empirical Models of Auctions. NBER Working Paper #12126. Available at:

Atherly, A., Dowd, B.E., and Feldman, R. (2004). The effect of benefits, premiums, and health risk on health plan choice in the Medicare program. Health Services Research, 39 (4), 847-864.

Bajari, P., and Tadelis, S. (2001). Incentives versus transaction costs: a theory of procurement contracts. RAND Journal of Economics, 32 (3), 387-407.

Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2011). How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program. NBER WP #16977. Available at: http://www.nber.org/papers/w16977.

Cawley, J.H., Chernew, M.E., and McLaughlin, C. (2002). HCFA payments necessary to support HMO participation in Medicare managed care. In: Garber, A.M. (Ed.), Frontiers of Health Policy Research, vol. 5. NBER, Cambridge, MA.

Cawley, J., Chernew, M., and McLaughlin, C. (2005). HMO participation in Medicare+Choice. Journal of Economics and Management Strategy, 14 (3), 543-574.

Chernew, M., DeCicca, P., and Town, R., (2008). Managed care and medical expenditures of Medicare beneficiaries. Journal of Health Economics, 27 (6), 1451-61.

Dowd, B. E., Feldman, R., and Coulam, R. (2003). The Effect of Health Plan Characteristics on Medicare + Choice Enrollment. Health Services Research, 38, 113-135.

Feldman R., Coulam R., and Dowd B. (2012) Competitive bidding can help solve Medicare's fiscal crisis. Washington, DC: American Enterprise Institute.

Gold, M. (2012). Medicare Advantage - Lessons for Medicare's Future. New England Journal of Medicine, Feb 22.

Gowrisankaran, G. and Town, R. (2006). Managed Care, Drug Benefits and Mortality: An Analysis of the Elderly, NBER WP #10204. Available at: http://www.nber.org/papers/w10204.

Guram, J.S. and Moffit, R. E. (2012). The Medicare Advantage Success Story — Looking beyond the Cost Difference. New England Journal of Medicine, Feb 22.

Hansen, R. G. (1988). Auctions with endogenous quantity. RAND Journal of Economics, 19 (1), 44-58.

Hendricks, K., and Porter, R. H. (2007). An Empirical Perspective on Auctions. In Handbook of Industrial Organization, M. Armstrong and R. Porter, eds., 3 (6), 2073-2143.

Klemperer, P. (1999). Auction Theory: A Guide to the Literature. Journal of Economic Surveys, 13 (3), 227-286.

Laffont, J. J., and Tirole, J. (1993). A Theory of Incentives in Procurement and Regulation. Cambridge, Mass.: MIT Press.

Maruyama, S. (2011). Socially optimal subsidies for entry: the case of Medicare payments to HMOs. International Economic Review, 52 (1), 105-129.

McAfee, R. P., and McMillan, J. (1987). Auctions and Bidding. Journal of Economic Literature, 25, 699-738.

McGuire, T.G., Newhouse, J.P., and Sinaiko, A.D. (2011). An Economic History of Medicare Part C. The Milbank Quarterly, 89 (2), 289-332.

Medicare Payment Advisory Commission (2010). Report to Congress: Medicare Payment Policy, March, Washington, DC.

Newhouse, J.P. (1996) Reimbursing health plans and providers: Efficiency in production versus selection. Journal of Economic Literature, 34, 1236-1263.

Office of the Massachusetts Attorney General Martha Coakley (2011). Examination of Health Care Cost Trends and Cost Drivers: Report for Annual Public Hearing, June 22, 2011. (http://www.mass.gov/ago/docs/healthcare/2011-hcctd.pdf)

Pizer, S. and Frakt, A. (2002). Payment policy and competition in the Medicare + Choice Program. Health Care Financing Review, 24 (1), 83–94.

Town, R. and Liu, S. (2003). The welfare impact of Medicare HMOs. RAND Journal of Economics, 34 (4), 719–736.

Vickrey, W (1961). Counterspeculation, Auctions, and Competitive Sealed Tenders. Journal of Finance, 16, 8-37.

Wilensky, G.R. (2012). Directions for Bipartisan Medicare Reform. New England Journal of Medicine, 366, 1071-1073.

Wooldridge, J.M. (2002). Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, MA.

Wyden, R., Ryan, P (2011). Guaranteed choices to strengthen Medicare and health security for all: bipartisan options for the future. (http://budget.house.gov/UploadedFiles/WydenRyan.pdf).

3.9. Supporting Materials

I repeat the analysis by adding LPPO plans to the sample, as well as among all plans by additionally including PFFS plans. In general, I find smaller effects with the inclusion of less managed plans. With LPPOs included, I find that bids increase $0.47 for every dollar increase in the benchmark (Section 3.9.1). Note that the number of markets increases. This is because in many (rural) counties, there can be a lack of HMO plans. Local PPOs and often private fee-for-service (PFFS) plans are serving such counties.

With LPPOs and PFFS included, bids increase $0.34 for a dollar increase in the benchmark (Section 3.9.2). This suggests that PFFS plans may be more responsive to competition than HMO and LPPO plans, which is also consistent with the increasingly negative coefficients on "number of insurers" from the HMO-only sample (Table 3.1) to the sample with LPPO and PFFS plans included (Section 3.9.2). HMO, LPPO, and PFFS plans all compete with each other simultaneously, as they face the same benchmarks in all counties and enrollees are free to choose among them. Rural counties are more likely to have a dominance of PFFS plans, as HMO plans are largely confined to densely-populated areas.

Section 3.9.1.

Table 3.4. Effect of benchmark on bids, market-level longitudinal model (HMO and local PPO plans).*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
VARIABLES	Base	No risk	No plan risk	No FFS	Linear year	Unweighted	Large markets	Interactions	HHI	HHI*bench
Benchmark	0.466***	0.464***	0.456***	0.525***	0.536***	0.420***	0.488***	0.567***	0.441***	0.501***
	(0.0680)	(0.0690)	(0.0688)	(0.0507)	(0.0505)	(0.0715)	(0.0627)	(0.0635)	(0.0652)	(0.0861)
Risk (FFS)	-14.37		-17.72	-11.47	-34.45**	9.333	-17.09	-14.28	-12.70	-12.65
	(15.61)		(16.81)	(15.14)	(14.97)	(17.37)	(15.97)	(15.31)	(15.72)	(15.68)
Risk (Plan)	-41.04	-44.99*		-35.21	-56.40**	-51.00**	-38.62	-45.25*	-40.04	-38.93
	(24.58)	(26.10)		(25.83)	(23.19)	(20.26)	(26.34)	(24.08)	(24.71)	(24.39)
FFS cost	0.0887*	0.0844*	0.0814*		0.0742	0.141**	0.0725	0.106**	0.0900*	0.0875*
	(0.0474)	(0.0466)	(0.0476)		(0.0465)	(0.0553)	(0.0469)	(0.0509)	(0.0453)	(0.0443)
Pop (1000s)	0.00770	0.00972	0.00749	0.0114	0.0131	0.0615	0.00214	0.0125	-0.0106	-0.0135
	(0.0447)	(0.0451)	(0.0443)	(0.0466)	(0.0460)	(0.0420)	(0.0446)	(0.0446)	(0.0491)	(0.0481)
Insurers	-1.451***	-1.459***	-1.500***	-1.357***	-1.798***	-2.087***	-1.336***	2.434	-1.287***	-1.312***
	(0.307)	(0.312)	(0.299)	(0.276)	(0.342)	(0.374)	(0.300)	(1.814)	(0.333)	(0.359)
Insurers*Bench								-0.00449**		
								(0.00213)		
HHI (beds)									-0.00737**	0.0211
									(0.00299)	(0.0257)
HHI*Bench										-3.23e-05
										(2.70e-05)
Market FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y		Y	Y	Y	Y	Y
Year (linear)					3.505**					
					(1.410)					
Constant	338.8***	332.7***	316.8***	344.0***	-6,699**	327.2***	330.2***	243.1***	371.2***	319.2***
	(43.91)	(45.04)	(44.61)	(38.34)	(2,829)	(46.32)	(41.12)	(48.46)	(43.62)	(72.70)
Observations	7,056	7,056	7,056	7,056	7,056	7,056	4,621	7,056	7,056	7,056
R-squared	0.381	0.379	0.377	0.375	0.346	0.332	0.396	0.385	0.386	0.387
Markets	2,231	2,231	2,231	2,231	2,231	2,231	1,484	2,231	2,231	2,231

Section 3.9.2.

Table 3.5. Effect of benchmark on bids, market-level longitudinal model (HMO, local PPO, and PFFS plans).*

VARIABLES	(1) Base	(2) No risk	(3) No plan risk	(4) No FFS	(5) Linear year	(6) Unweighted	(7) Large markets	(8) Interactions	(9) HHI	(10) HHI*bench
Benchmark	0.338***	0.334***	0.344***	0.513***	0.629***	0.329***	0.362***	0.389***	0.338***	0.285
	(0.0664)	(0.0646)	(0.0654)	(0.0526)	(0.0617)	(0.0597)	(0.0793)	(0.0744)	(0.0660)	(0.178)
Risk (FFS)	-23.64*		-20.61	-24.58	-28.90*	-9.423	-32.10***	-23.23	-23.61	-23.92
	(14.06)		(13.94)	(16.69)	(15.82)	(10.69)	(11.85)	(14.23)	(14.16)	(14.39)
Risk (Plan)	55.14	52.15		71.80**	71.03**	57.72**	46.83	49.94	54.98	54.66
	(34.95)	(33.92)		(35.19)	(34.01)	(22.17)	(52.78)	(37.76)	(35.27)	(34.76)
FFS cost	0.285***	0.286***	0.301***		0.107	0.350***	0.217**	0.276***	0.286***	0.284***
	(0.0750)	(0.0731)	(0.0707)		(0.0679)	(0.0656)	(0.0902)	(0.0750)	(0.0756)	(0.0772)
Pop (1000s)	-0.0468	-0.0429	-0.0440	-0.0256	-0.0249	0.0273	-0.0483	-0.0240	-0.0449	-0.0386
	(0.0835)	(0.0784)	(0.0833)	(0.0815)	(0.0977)	(0.0603)	(0.0858)	(0.0976)	(0.0867)	(0.0853)
Insurers	-1.968***	-1.987***	-1.898***	-1.934***	-2.759***	-2.867***	-1.722**	1.802	-1.981***	-2.000***
	(0.639)	(0.620)	(0.673)	(0.658)	(0.804)	(0.531)	(0.653)	(4.516)	(0.651)	(0.634)
Insurers*Bench								-0.00456		
								(0.00555)		
HHI (beds)									0.000819	-0.0205
									(0.00445)	(0.0632)
HHI*Bench										2.59e-05
										(7.56e-05)
Market FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y		Y	Y	Y	Y	Y
Year (linear)					11.32***					
					(1.960)					
Constant	247.9***	229.7***	279.0***	292.0***	-22,586***	196.4***	292.8***	216.0***	245.9***	291.0*
	(58.05)	(60.14)	(48.60)	(55.17)	(3,919)	(47.47)	(79.46)	(54.48)	(54.96)	(163.1)
Observations	12,334	12,334	12,334	12,334	12,334	12,334	5,846	12,334	12,334	12,334
R-squared	0.659	0.657	0.655	0.640	0.626	0.705	0.623	0.660	0.659	0.659
Markets	3,106	3,106	3,106	3,106	3,106	3,106	1,473	3,106	3,106	3,106

Page intentionally left blank.